RF Energy Harvesting for INTERNET OF THINGS

K.J. Vinoy

ECE Dept
Indian Institute of Science
Bangalore 560012

ECE Dept, Indian Institute of Science

- **Vision Statement:** Excellence in Theoretical and Experimental Research in Communications, Signal Processing, Microelectronics and RF/Photonics.

- **Faculty: 24;** Fellows of IEEE: 4; Fellows of INAE: 8

- Active in Publications: Books, Book Chapters, Journal & Conference Papers; Patents; Standardization etc
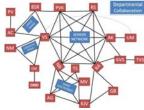
- Collaborative research

People
Masters Students
[ME, MSc→ Mtech, Mtech(Res)]
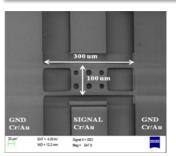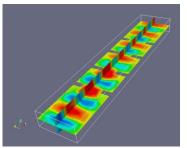PhD Students
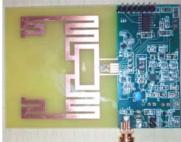Project Staff

ECE: Microwave Engineering

K.J. Vinoy

- Low-Actuation Voltage Capacitive RF MEMS Switch (<10V)
  - Low-complexity fabrication process to enhance process yield
  - High reliability: no failure even up to 10 million cycles of operation tested
- Meso-scale Electrostatic Phase Shifter on microwave Laminate (MEPL)
  - Utilizes modern printed circuit board fabrication technology.
  - X-band monolithic antenna array system on the microwave laminate board demonstrated.

- Wideband group delay engineering in RF circuits for radar, medical imaging, and spectrum sensing.
  - Demonstration uses two stage All – Pass Networks; can be extended over multiple stages to obtain a higher bandwidth and/or higher group delay slope.
- RF energy harvesting circuits
  - Integrated with RF transmitters and sensors for practical IoT nodes
  - High efficiency RF-DC converter which can operate at input power of -20dBm (10µW) at 2.4GHz using UMC 130nm process MOSFETs.
- FEM based algorithms for Electromagnetic circuits & components (periodic structures such as metamaterials)
  - Fast computation of electromagnetic propagation characteristics
  - Especially suited for evaluation of processes uncertainties

- **Introduction**
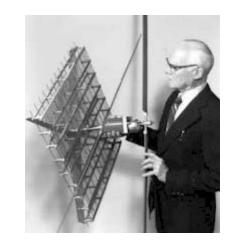 - **Wireless Power Transfer**
    - **Energy Harvesting**
  - **Internet of Things**

- Highlights of Recent Development (Hardware)
  - Powering wireless terminals
  - RFID with integrated sensors

  Ongoing Research Challenges

Wireless Power Transfer (WPT)

- Indicates transfer of electric energy remotely

- WPT has a long history!!
  - Tesla demonstrated it in 1899 by wirelessly powering fluorescent lamps 40 kms away from the power source.
  - Had multiple patents in early 1900s.

- In 1960s W.C. Brown coined the term Rectenna, which he used to directly converts incoming microwaves to DC.
  - He demonstrated its ability to power a helicopter solely through microwaves for 10 hours continuously.
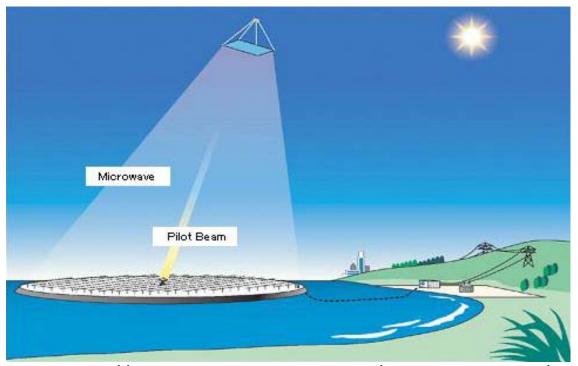
- These demonstrations involved dedicated sources with **large power to transmit over long distances**.

http://mainland.cctt.org/istf2008/brown.asp

SSPS

- Space Solar Power Satellites
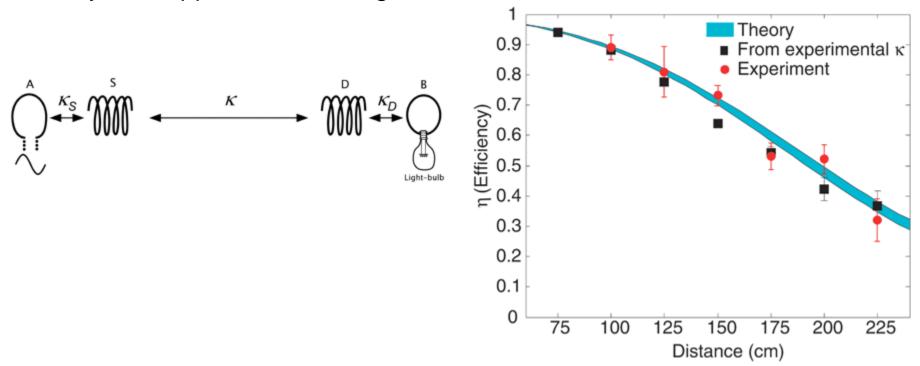
- WPT is widely investigated for putting **solar power generating satellites** into space and transmitting power to Earth stations. (Mainly in Japan)

http://www.jspacesystems.or.jp/en_project_ssps/
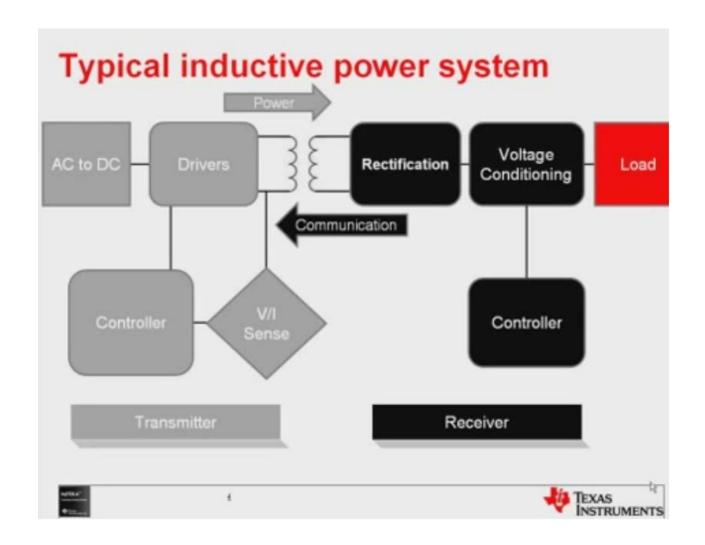
Near field Wireless Power Transfer

- Recent demonstration by MIT to transfer high RF power (Watts) transferred across meters.

- Resonant coils are used

- Typically at 100 kHz to 10's of MHz

- Many new applications emerged

André Kurs et al, Wireless Power Transfer via Strongly Coupled Magnetic Resonances , Science, Vol. 317 no. 5834 pp. 83-86. 6 July 2007.

Source: Texas Instruments Qi Development kit
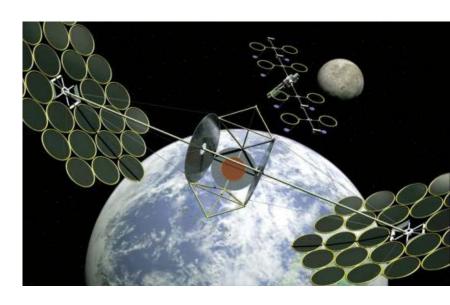
Far → Near in WPT
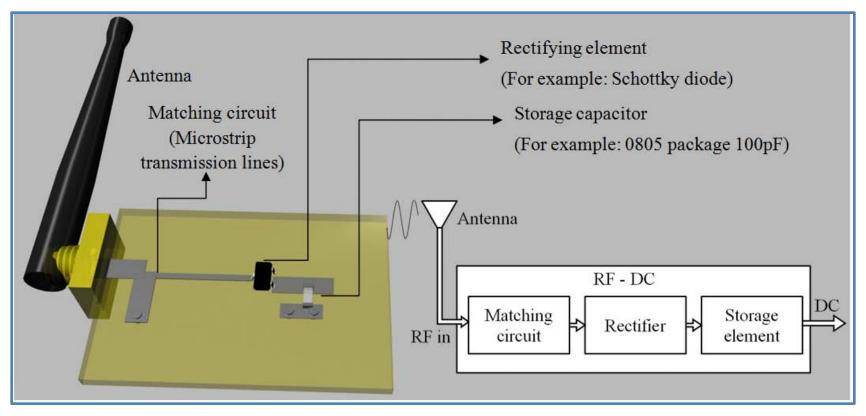
- Free space loss factor is a major bottleneck for power transfer at large distances

- Short distance/ Near field options
  - MIT demonstration (2007)
  - Qi Standard
  - Phone charging solutions
  - Vehicles running on wireless power

- Two extremes in WPT
  - mW ← → MW
  - mm ← → 1000s km
  - 100kHz ← → 2.4/5.8GHz
  - 10cm x 10cm ← → km x km
  - Commercial vs bluesky

Far Field Transfer of RF Energy

- Focus of this talk

- Applications: RFID tags, Wireless Sensor Network nodes, biomedical equipment, home automation and structural monitoring can benefit from RF energy harvesting.

- Block diagram and a design example:

A New Paradigm: Internet of Things (IoT)

- IoT refers to uniquely identifiable objects and their virtual representations in an Internet-like structure.
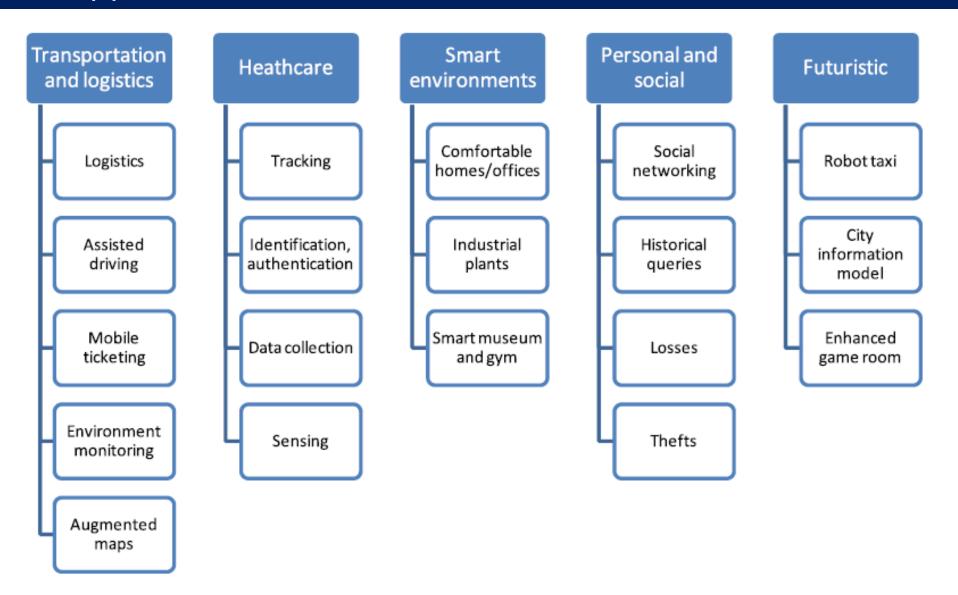
> Connects Anytime, Anyplace for Anyone (ICT)
>
> AAA + for Anything   (IoT)

- IoT is a scheme for connecting things: sensors, actuators, and other smart technologies, thus enabling person-to-object and object-to-object communications.

- **Continuous availability of power is crucial for their deployment**

11

IoT Applications

Transportation and logistics
- Logistics
- Assisted driving
- Mobile ticketing
- Environment monitoring
- Augmented maps

Heathcare
- Tracking
- Identification, authentication
- Data collection
- Sensing

Smart environments
- Comfortable homes/offices
- Industrial plants
- Smart museum and gym

Personal and social
- Social networking
- Historical queries
- Losses
- Thefts

Futuristic
- Robot taxi
- City information model
- Enhanced game room

L. Atzori , A. Iera, G. Morabito, The Internet of Things: A survey Computer Networks 54 (2010) 2787–2805

Comparison of different wireless protocols

Today, a lot can be done at low power!!

Characteristics of key 2.4GHz ISM Band Radios studied :

	BLE	ANT	Zigbee	WLAN
Topologies	P2P , Star	P2P , Star, tree, mesh	P2P , Star, mesh	P2P , Star
Modulation	GFSK	GFSK	OQPSK	DSSS (802.11b)
Max data rate	1Mbps	12.8-60 Kbps	250Kbs (@2.4Ghz)	1- 11Mbps (802.11b)
Throughput	305 kbps	20Kbps	100Kbps	6Mbps (802.11b)
Range (in m)	10-100(0-10dBm)	30 (@ 0dBm)	10-100 (0-20dBm)	100+(20dBm)
Max nodes in piconet	7	65533	Star-65536	32-64
Battery life	1-2 years (coin cell)	1-2years (coin cell)	100-1000 days	0.5-5days

Key Aspects :

BLE is robust and has lowest power consumption but cannot natively form mesh networks,

Zigbee can support large mesh networks, power consumption is higher than BLE and throughput is lower: It is suitable for low data rate, low power, large size networks.

WLAN is primarily suitable for transferring bulk data at high speeds, Not suitable for low power applications.

Power Requirements in Common WSN

	Crossbow MICAz	Intel IMote2	Jennic JN5139
Radio standard	IEEE802.15.4/ZigBee	IEEE802.15.4	IEEE802.15.4/ZigBee
Typical range	100m (outdoor), 30m (indoor)	30m	1 km
Data rate (kbps)	250 kbps	250 kbps	250 kbps
Sleep mode (deep sleep)	15 μA	390 μA	2.8 μA (1.6μA)
Processor only	8mA active mode	31–53mA*	2.7+0.325mA/MHz
RX	19.7mA	44mA	34mA
TX	17.4mA (+0dbm)	44mA	34mA (+3 dBm)
Supply voltage (minimum)	2.7V	3.2V	2.7V
Average	2.8mW	12mW	3mW

JM. Gilbert∗ F. Balouchi, Comparison of Energy Harvesting Systems for Wireless Sensor Networks, International Journal of Automation and Computing 05(4), October 2008, 334-347

	Jennic JN5148	TI- CC430	BLE	Zarlink ZL70250
Active mode current at 16MHz [mA]	6	4	6.7	3.2
Deep sleep current [nA]	100	1000	400	20
Transmission current [mA]@Tx-power [dBm]	15@2.5	18@0	36@2	2@-10
Transmit frequency	2.4 GHz	2.4 GHz	2.4 GHz	868 MHz
Wakeup time [ms]	1	3	0.12	0.16
Energy consumption for a transmission cycle of 2ms [μJ]	183	300	196	32
Power supply voltage [V]	2.2 – 3.6	1.8 -3.6	2-3.6	1.2 – 1.8

In perspective

- Energy requirements in different devices/systems

- 6 orders of magnitude variation!!!

- Energy requirements in WSN

  – Depends on the complexity/ standard/ range

  – eg 90 μW to power a pulse oxymeter sensor, to process data and to transmit them at intervals of 15 s

Device type	Power consumption
Smartphone	1W
MP3 decoder chip	58 mW
Hearing aid	1 mW
Wireless sensor node	100 μW*
RF receiver chip	24 mW
GPS receiver chip	15 mW
6D motion sensor	14.4 mW
Cell phone (standby)	8.1 mW
PPG sensor	1.473 mW
Humidity	1 mW
Pressure	0.5 mW
3D accelerometer	0.324 mW
Temperature	27 µW
Cardiac pacemaker	50 µW
Wristwatch	7 µW
Memory R/W	2.17 µW
A-D conversion	1 µW

J Yun, S. Patel, M.Reynolds, G. Abowd "A Quantitative Investigation of Inertial Power Harvesting for Human-powered Devices," UbiComp'08, September 21-24, 2008, Seoul, Korea.

R.J.M. Vullers, et.al, Micropower energy harvesting, Solid-State Electronics 53 (2009) 684–693

15

Power Density from Various Harvesters

Ambient RF	< 1 µW/cm^2
Ambient light	100 mW/cm^2 (directed toward bright sun)
	100 µW/cm^2 (illuminated office)
Thermoelectric	60 µW/cm^2
Vibrational	4 µW/cm^3 (human motion ~Hz)
microgenerators	800 µW/cm^3 (machines ~kHz)
Ambient airflow	1 mW/cm^2
Push buttons	50 µJ/N
Hand generators	30 W/kg
Heel strike	Up tp 7 W for 1 cm deflection

Power requirements in conventional sensor network nodes may not be met by harvesting alone!

J Yun, S. Patel, M.Reynolds, G. Abowd "A Quantitative Investigation of Inertial Power Harvesting for Human-powered Devices," UbiComp'08, September 21-24, 2008, Seoul, Korea.

Electromagnetic: Solar → RF

- PV is a good source of energy. Recall,

Ambient light	100 mW/cm^2 (directed toward bright sun)
	100 μW/cm^2 (illuminated office)

- However, it light is not always available.

  – At night

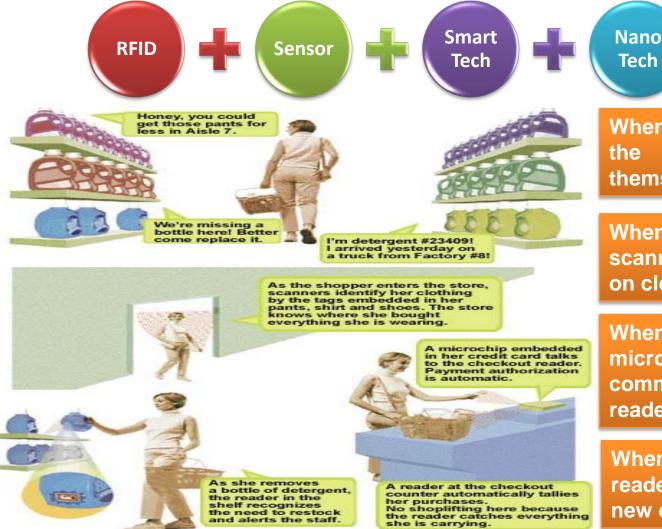 – Specific scenarios: in a closed chamber, or mine.

- RF is an alternative

  – Unlike others, RF sources may be ambient or intentional.

  – Ambient sources such as base stations or broadcast stations

  – Special sources: RF ID reader, Phone charger, special beacons

Demand and Supply

- The peak currents needed during transmit and receive operation is not achievable using the harvester alone.

- Buffering is also needed to ensure continuous operation during times without power generation.

- The combination of an energy harvester with a small-sized storage is the best approach to enable energy autonomy of the network over the entire lifetime.

  - Rechargeable battery

  - Thin film batteries
    - can be integrated directly in Integrated Circuit (IC) packages in any shape or size,
    - Flexible when fabricated on thin plastics
    - Thin film batteries have high impedance;
    - Low discharge efficiency compared to Li-ion batteries

  - super capacitor
    - Leakage in super capcitors depends on the voltage. Low at low voltage

Internet of Things (IoT)

Embedding short-range mobile transceivers into a wide array of gadgets and everyday items, enabling new forms of communication between people and things, and between things themselves.

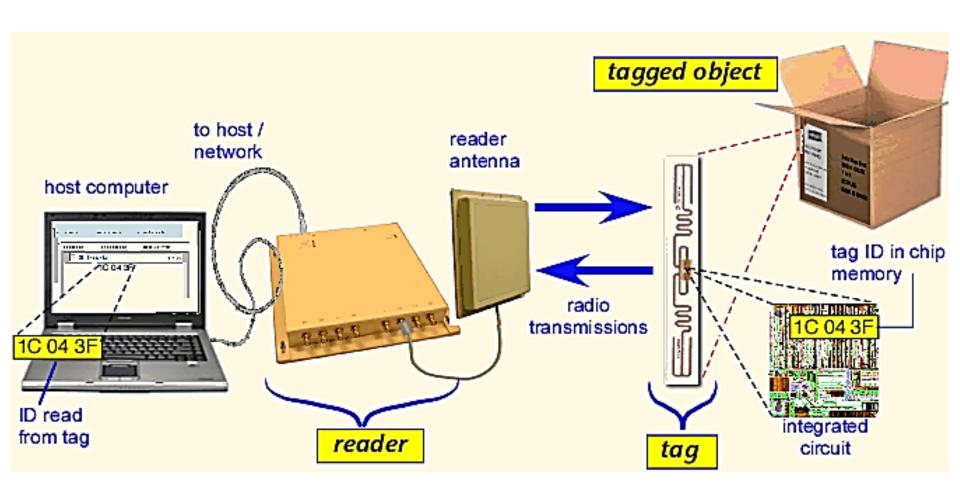When shopping in the market, the goods will introduce themselves.

When entering the doors, scanners will identify the tags on clothing.

When paying for the goods, the microchip of the credit card will communicate with checkout reader.

When moving the goods, the reader will tell the staff to put a new one.

Introduction to RFID

- The reader converts incident field and returns useful data

- In passive RFID systems reader transmits EM energy that "wakes up" the tag and provides power for the tag to respond to the reader.

tagged object

to host / network

reader antenna

host computer

tag ID in chip memory

1C 04 3F

radio transmissions

1C 04 3F

ID read from tag

reader

tag

integrated circuit

- **Backscatter** is the reflection of signals back towards their source.

  – In this scheme, two devices communicate using incident (or ambient) RF as the source of power.

  – Backscattering is achieved by changing the impedance of a receiver in the presence of an incident signal.

  – When waves encounter a new media that have different impedances, a part of the wave is reflected.

  – The reflection depends on the difference in the impedances.

- By modulating the impedance at the receiver port, one can control the scattered RF energy, hence enabling information transmission.
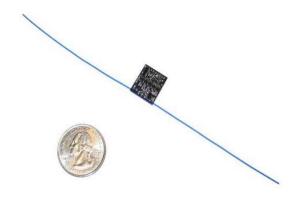
abc.cs.washington.edu/files/comm153-liu.pdf
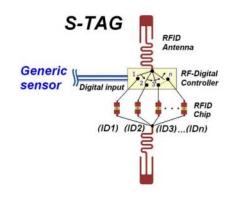
RFID → IoT

- RFID

  – Uses radio waves for identifying or tracking the object.

  – Proven to be a simple and cost effective system

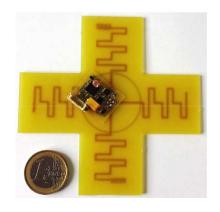 – Tags are very cheap and is possible to be attached to everyday objects.

- RFID is considered a prerequisite of **Internet of Things.**

- **Example: RFID tags can be integrated with sensors**

  – When a reader reads a tag, the sensor information will be sent to the reader along with the identity of the object.
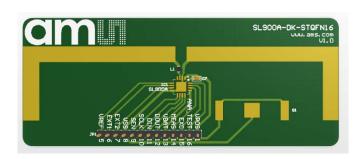
Some Examples

Discrete Element Based
WISP (Wireless Identification Sensing Platform)

Multi-Chip Based S-tag

Chips with I2C / SPI
SPARTACUS / RAMSES (Self-Powered Augmented RFID Tag for Autonomous Computing and Ubiquitous Sensing / RFID Augmented Module for Smart Environmental Sensing)

Single IC Based Sensing Tags
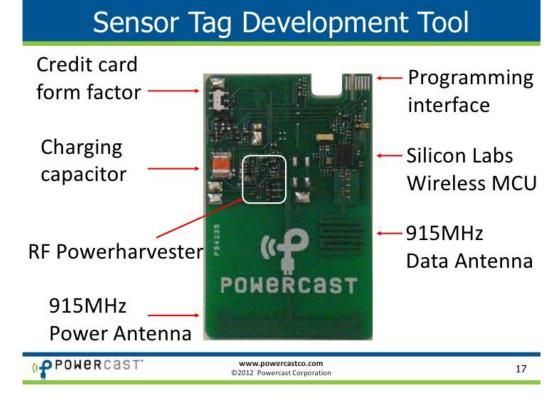

Printed Chipless RFID Tag

WISP

- Wireless Identification Sensor Platform (2009)

  – WISPs are a wireless, battery-free sensing and computation platform, powered by harvested energy from off-the-shelf UHF RFID readers.

  – To a RFID reader, a WISP is a EPC gen1 or gen2 tag; but inside the WISP, the harvested energy is operating a 16-bit general purpose microcontroller.

  – The microcontroller can perform computing tasks, including sampling sensors, and communicate to the RFID reader.

  – WISPs have been built with various sensors, WISPs can write to flash and perform cryptographic computations.

A collaboration between Intel Research Seattle and the University of Washington.

http://wisp.wikispaces.com/Wisp+4.1+DL

RFID Sensors (Products)

- ID operation is passive; yet most sensors require power sources

- Powercast has a wireless sensor that is battery-less. Uses RF energy harvesting.

- Harvesting schemes works at power as low as -12dBm. (RF-DC conversion efficiency above 40% only above -8dBm)

- Harvested power >0.4mW for RF in of >-1dBm.

- Multiple custom ICs and discretes

- Other suppliers include
  - Phase IV
  - RFID sensor systems
  - etc

Sensor Tag Development Tool

Credit card form factor →

Charging capacitor →

RF Powerharvester →

915MHz Power Antenna →

→ Programming interface

← Silicon Labs Wireless MCU

← 915MHz Data Antenna

PS4235

POWERCAST

www.powercastco.com
©2012 Powercast Corporation

17

http://www.powercastco.com/uhf-rfid-sensing-passive-rfid-wireless-sensor-tags/
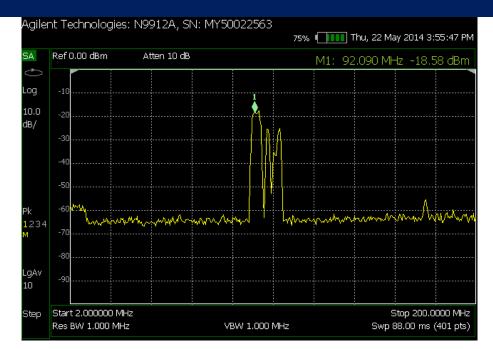
Battery-less Wireless Terminals

- Most of our work in this direction was towards battery-less terminals

- Long life terminals without wiring

- These are useful when

  - Terminals are embedded within structures (or body)

  - Devices to be deployed in hostile environments

  - Use of battery is not allowed (potential cause for explosion)

- Other factors

  - Cost, weight, etc.

- Primary focus: use of radio frequencies (ambient/intentional)

Ambient RF Sources

- Several sources:

  - WiFi Access points (mW) [2.4/.6GHz]

  - Cellular Tower (W) [900/1800 MHz]

  - TV Broadcast (MW) [150-450MHz]

  - FM broadcast (kW) [90-108MHz]

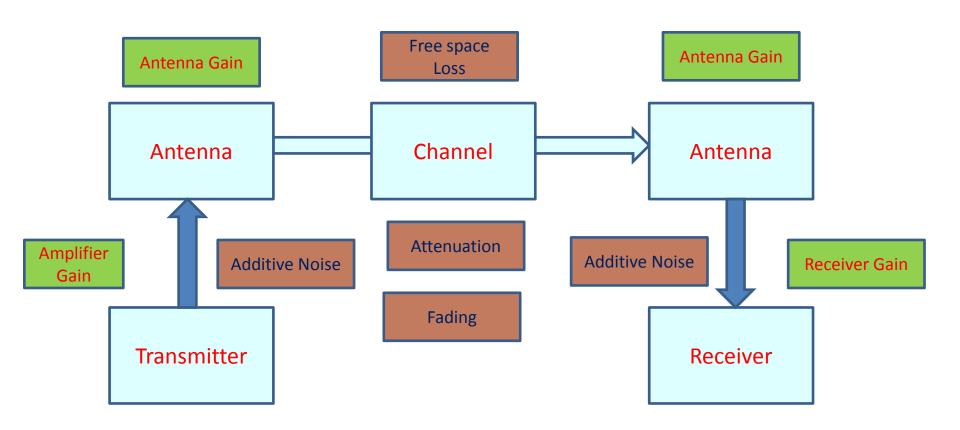 - AM Radio broadcast (kW) [<1MHz]

- In general

  - Lower frequencies help non-line-of sight propagation

  - Power availability from ambient sources is limited and varies from place to place.

- Note

  - Unlike other sources, most practical RF harvesters (eg in RF ID) depend on **intentionally generated energy**.

  - This is called **wireless power transfer (WPT)** in the conventional RF/Microwave parlance.

Wireless Communication System

- Power transfer scheme is no different!!

Antenna Fundamentals: Directivity

*This is the ratio of the radiation intensity **in a given direction** to the radiation intensity averaged over all direction*

Average radiation intensity, $\quad U_0 = \dfrac{P_{rad}}{4\pi}$

Directivity,

$$D(\theta, \phi) = \frac{U(\theta, \phi)}{U_0} = \frac{4\pi U(\theta, \phi)}{P_{rad}}$$

➤ If direction is not specified, it implies the direction of maximum radiation intensity

$$D_{\max} = \frac{4\pi U_{\max}}{P_{rad}} \qquad\qquad D_{dB} = 10 \log D$$

Maximum directivity and Maximum effective area

The radiated power density by a transmitter at a distance R
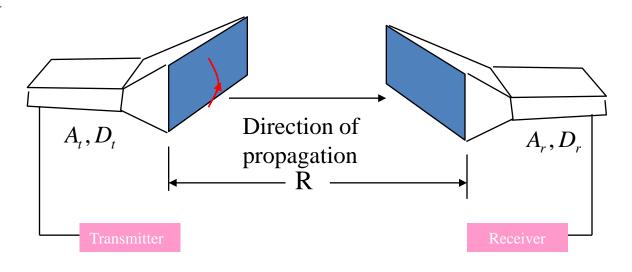
$$W_t = \frac{P_t D_t}{4\pi R^2}$$

Power received

$$P_r = W_t . A_r = \frac{P_t D_t A_r}{4\pi R^2}$$

$$or, \ D_t . A_r = \frac{P_r}{P_t} . 4\pi R^2$$

A_t, D_t

Direction of propagation
R

A_r, D_r

Transmitter

Receiver

$$\therefore \frac{D_t}{A_t} = \frac{D_r}{A_r}$$

$if, \ D_0 = \max. \text{directivity}$

$A_{em} = \max. \text{effective area}$

By reversing the transmission direction

$$D_r A_t = \frac{P_r}{P_t} \cdot 4\pi R^2$$

this can be generalized by

$$\frac{D_1}{A_1} = \frac{D_2}{A_2} = k = \frac{4\pi}{\lambda^2}$$
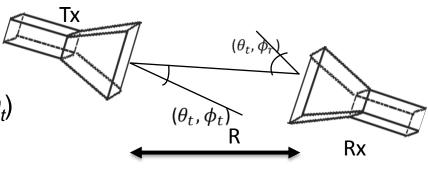
$$D = \frac{4\pi A_{em}}{\lambda^2}$$

Frii's Transmission Equation

- The radiation intensity for an isotropic radiator is $W_0 = \dfrac{P_t}{4\pi R^2}$

- For an antenna of gain G_t (or directivity D_t)

$$W_t = \frac{P_t G_t(\theta_t, \phi_t)}{4\pi R^2} = e_t \frac{P_t D_t(\theta_t, \phi_t)}{4\pi R^2}$$

- The effective aperture of a receiving antenna is given by $A_r = e_r D_r(\theta_r, \phi_r)\dfrac{\lambda^2}{4\pi}$

- Therefore,

$$P_r = e_r D_r(\theta_r, \phi_r)\frac{\lambda^2}{4\pi} W_t = e_t\, e_r D_t(\theta_t, \phi_t) D_r(\theta_r, \phi_r)\frac{\lambda^2}{(4\pi R)^2} |\hat{\boldsymbol{\rho}}_t \cdot \hat{\boldsymbol{\rho}}_r|^2$$

$$\frac{P_r}{P_t} = e_t e_r D_t(\theta_t, \phi_t) D_r(\theta_r, \phi_r)\frac{\lambda^2}{(4\pi R)^2} |\hat{\boldsymbol{\rho}}_t \cdot \hat{\boldsymbol{\rho}}_r|^2$$

$$\frac{P_r}{P_t} = e_{cdt} e_{cdr}(1 - |\Gamma_t|^2)(1 - |\Gamma_r|^2) D_t(\theta_t, \phi_t) D_r(\theta_r, \phi_r)\frac{\lambda^2}{(4\pi R)^2} |\hat{\boldsymbol{\rho}}_t \cdot \hat{\boldsymbol{\rho}}_r|^2$$

- When the antennas are pointing towards each others' peak radiation direction,

$$\frac{P_r}{P_t} = G_{0t} G_{0r}\left(\frac{\lambda}{4\pi R}\right)^2$$

Note that includes a loss factor (usually called **Free space Loss factor)**
Does not include dissipation/attenuation in medium; caused by spreading

Some numbers on Radiative form of WPT…

- Practical systems will have

  - Operational frequencies in ISM bands.

  - Most terminals are compact.

  - Antenna efficiency is compromised.

  - Nearly isotropic radiations expected.

- Main bottleneck is the physical limits in transmission.

$$P_r = P_t * G_t * G_r * (\lambda/4\pi r)^2$$

  - At 1 GHz (λ=30cm) r=1m; Antenna gain @0dBm, free space loss factor is about **0.06%**

  - Even with a moderate gain transmitter antenna (6dBi) power received @1m for 1W transmission, is just 2mW.
    - Drops to 23μW at 10m !!
    - **The voltage of the signal is low!!**

- In radiative power transfer, Distance from transmitter is a major concern.
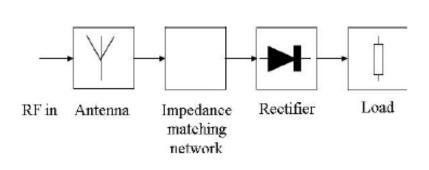
- **Harvesting of ambient radiations or Radiative transfer of energy addressed**

  - Is it possible to harvest the RF energy from base stations

  - Are there other viable sources of RF energy

- Can low power communication systems be designed to operate entirely from harvested energy

  - Integrate sensors, control, etc

- Can we use RF EH/ WPT to increase the range of backscatter communication (RFID scenario)
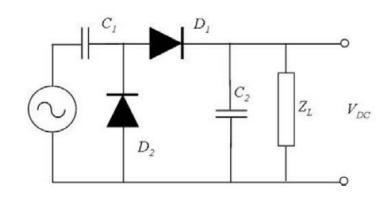
Design of Rectifiers

Required for converting incoming RF into DC power.

The challenge lies in maximizing the power conversion efficiency for low input power and minimizing the dimensions.

- RF to DC conversion by rectification of the incident RF signal by a **Schottky diode**

    – Most diodes have a finite cut-in voltage

    – Diode is a non-linear device (performance depends on current or load)

    – Impedance matching required between antenna and diode

    – In most cases, the input voltage needs to be boosted

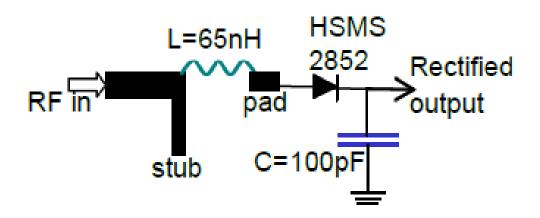Conceptual diode based Rectenna

Voltage doubler

Voltage magnification in Matching circuit

- Matching circuit is required to provide impedance match between antennas (50Ω typical) to diode terminated with high impedance load (capacitor and/or high R in parallel).

  - LC matching networks provide voltage magnification.

  - This helps the diode conduct a good fraction of half cycle.

- The higher the voltage across the diodes, the more efficient the rectifier gets.

  - In practice Q is limited

  - Applications requiring higher voltages, a voltage multiplier configuration is used.

$$V_C = V_{in} \times \frac{\dfrac{1}{j\omega C}}{R + j\omega L + \dfrac{1}{j\omega C}}$$

At resonance,

$$V_C = V_{in} \times \frac{1}{j\omega_0 CR} = -jQV_{in}$$

Tuned Rectifier at RF
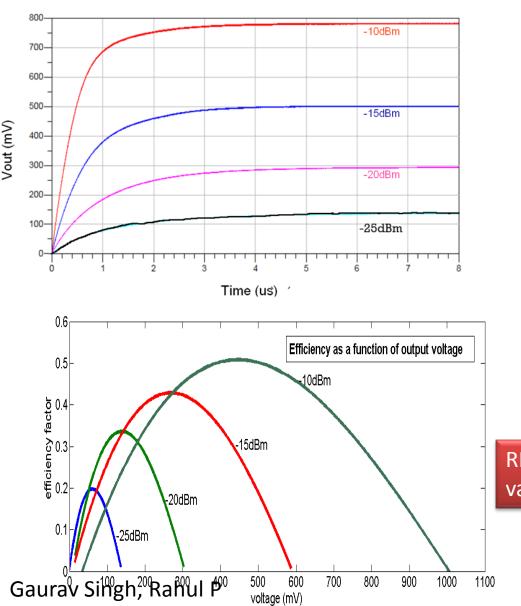
- A tuned rectifier implemented using discrete components
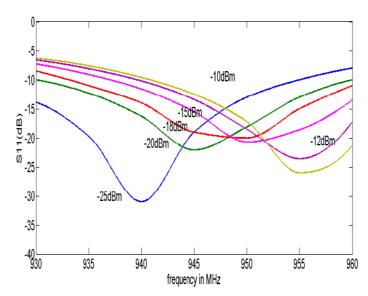
P_in→	-10dBm	-13dBm	-16dBm	-20dBm	-25dBm
Freq. ↓					
930MHz	917mV	664mV	469mV	281mV	131mV
945MHz	1016mV	736mV	515mV	300mV	132mV
955MHz	1038mV	747mV	513mV	289mV	122mV
960MHz	1032mV	736mV	499mV	276mV	114mV
Peak efficiency	51%	47%	39%	33%	20%

K. J. Vinoy, T. V. Prabhakar, A Universal Energy Harvesting Scheme for Operating Low-Power Wireless Sensor Nodes Using Multiple Energy Resources, pp. 453-466, Micro and Smart Devices and Systems, Springer 2014.

RF-DC Conversion efficiency depends on various conditions

Gaurav Singh, Rahul P

38
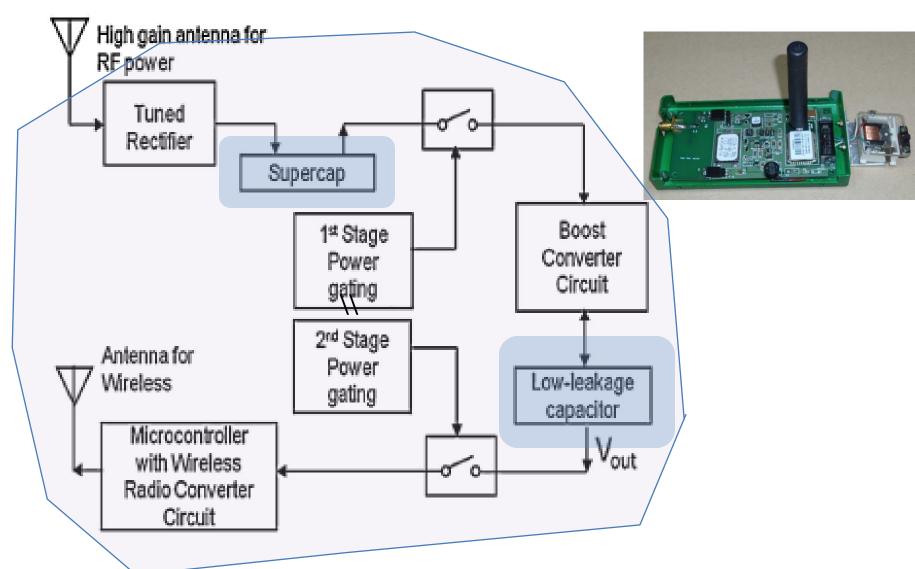
Rectifier Circuit using 4 diodes
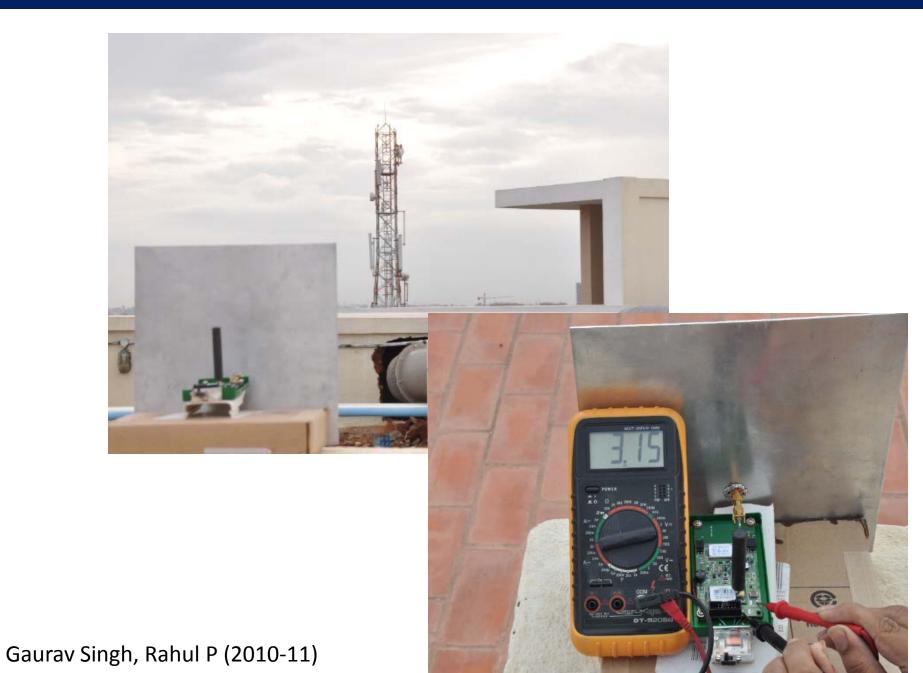
Power level (dBm)	Charging time (ms)	Efficiency (%)
0	40	64.11
-2	55	63.77
-3	67.5	63.5
-5	90	63
-7	230	59.89
-10	370	56.78
-12	500	53.89
-15	900	45
-18	2000	20.56

K. J. Vinoy, T. V. Prabhakar, A Universal Energy Harvesting Scheme for Operating Low-Power Wireless Sensor Nodes Using Multiple Energy Resources, pp. 453-466, Micro and Smart Devices and Systems, Springer 2014.
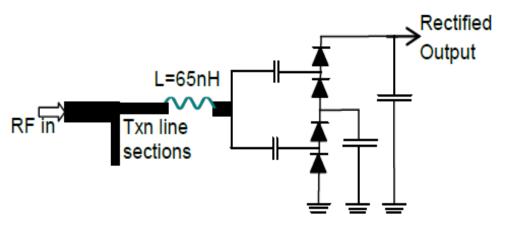
Aditya Mitra, Chaithanya (2011-12)

G. Singh, R. Ponnaganti, T. V. Prabhakar, and **K.J. Vinoy**, "A tuned rectifier for RF energy harvesting from ambient radiations," Int. J. Electronics & Communications, vol. 67, no. 7, pp. 564-569, July 2013

Gaurav Singh, Rahul P (2010-11)

Gaurav Singh, Rahul P (2010-11)

Characterization in Lab

- ## Using various antennas
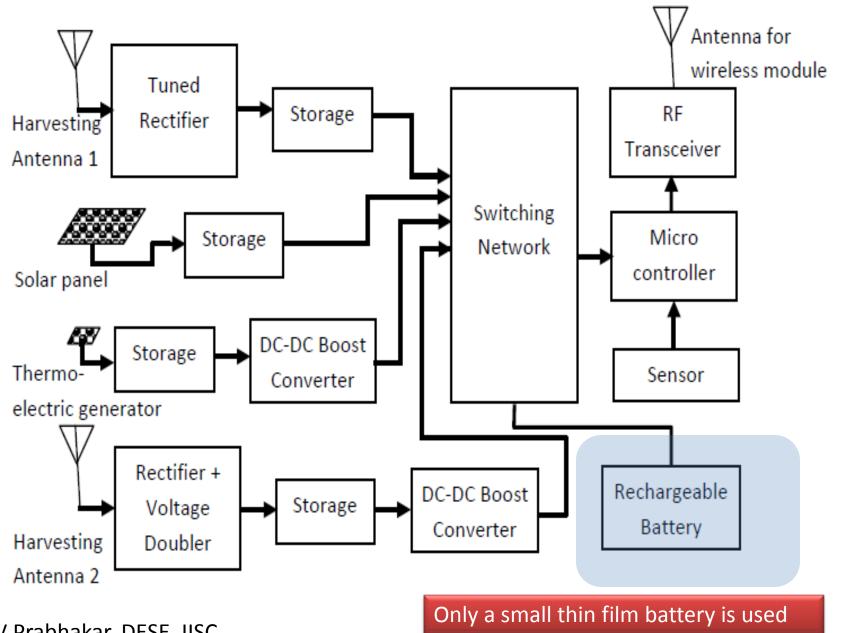
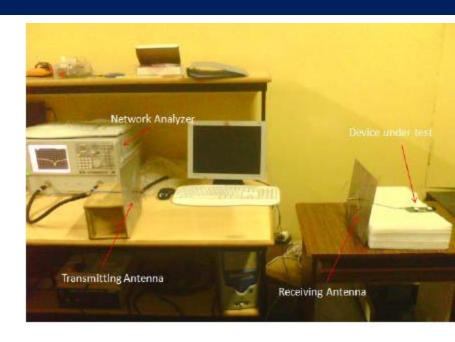| Distance from Transmitter [m]	1.5	2	2.5	3
Power received by dipole antenna [dBm]	-20.5	-22.1	-23.9	-25.2
Calculated power density [uW/cm^2]	0.078	0.055	0.035	0.03
Power received by patch antenna [dBm]	-15.1	-16.1	-17.6	-19.2
Transmit interval [mm:ss]	07:26	12:13	25:00	never
Power received by biquad antenna [dBm]	-11.8	-13.2	-14.9	-15.9
Transmit interval [mm:ss]	02:20	03:25	7:10	10:33

G. Singh, R. Ponnaganti, T. V. Prabhakar, and **K.J. Vinoy**, "A tuned rectifier for RF energy harvesting from ambient radiations," Int. J. Electronics & Communications, vol. 67, no. 7, pp. 564-569, July 2013

Gaurav Singh, Rahul P (2010-11)

Incident power above -18dBm required for operation

Only a small thin film battery is used

UEHP: Performance with different sources

Solar		RF		TEG	
Light Intensity (Lux)	Duty Cycle of operation (s)	Power Level (dBm)	Duty Cycle of Operation (s)	Temperature Differential (^{o}C)	Duty Cycle of Operation (s)
1000	7	0	3	55	9
300	11	-5	6	45	13
200	**20**	**-7**	20	35	240
100	42	-10	50	-	-
-	-	-12	240	-	-

An incident RF power of -7dBm (~0.2mW) performs similarly as at low light PV.

An appropriately oriented 20mW source with a high gain antenna (~10dB) can reach this RF power at a low gain rectenna (eg using PIFA) at 1 m distance.
Power levels within emission guidelines…

K. J. Vinoy, T. V. Prabhakar, A Universal Energy Harvesting Scheme for Operating Low-Power Wireless Sensor Nodes Using Multiple Energy Resources, pp. 453-466, Micro and Smart Devices and Systems, Springer 2014.

Aditya Mitra, Chaithanya (2011-12)

Other Possibilities using Wireless Power Transfer

- Power transfer by radiation is not efficient

- Waveguiding systems can ensure better transmission of power

  – Loss in waveguide is a small fraction of a dB/m (~0.2dB/m)

  – Metal ducts may carry higher order modes with higher losses

  – Extended to conducting ducts, Tunnels, mine shafts etc with some compromise

- Other possibilities

  – Surface wave

  – Focusing of fields

- Empty enclosures with metallic walls

  – Containers, tanks, airplane cabin, trains, etc

  – Other objects in the path may reduce the efficiency!

3: Antenna for RFID sensors
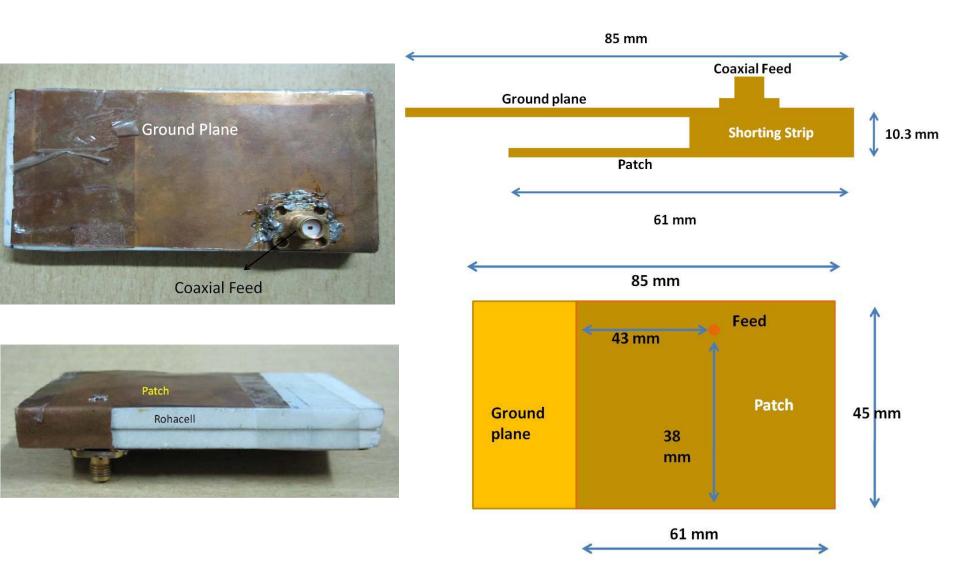
- Work involved design of antenna for RF harvesting sensors
    - These fuel level sensors to be deployed in a fuel tank of aircrft.
    - Optimization of design should focus on efficiency
    - High gain or directivity is not required.

- EH platform to be used with RFID sensors deployed inside fuel tank

- Requirements/Assumptions:
    - Incident energy is of random polarity and direction.
    - Operating frequency is 902MHz-928MHz.
    - Antenna must operate in air (relative permittivity = 1) and fluid (relative permittivity = 2.1)
    - Dimensions of planar antenna board:
        - Target dimensions: 3 in. x 2 in.
        - Maximum dimensions: 6 in. x 4 in.
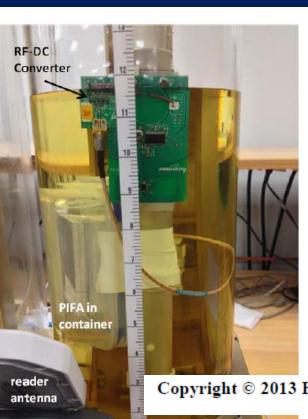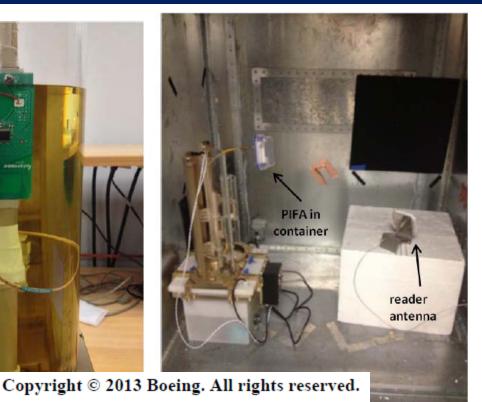
Antenna Design dimensions

Vivekanand M, Harikiran M

Measurements at Boeing (Nov 2013)

RF-DC Converter

PIFA in container

reader antenna

PIFA in container

reader antenna

After Integrating with Sensor and RFID board;
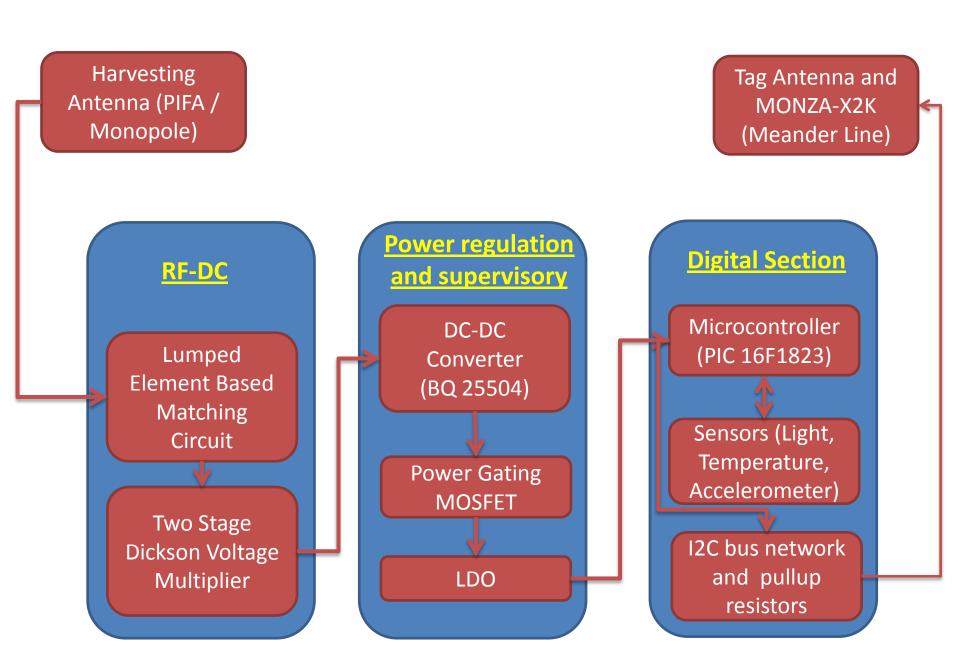
Measured in a room and reverberation chamber

Resolution of fluid height measurement to within 0.25".
1W maximum transmit power.
Uses a modified reader protocol.

ANRC
A BOEING IISC WIPRO HCL RESEARCH INITIATIVE IN INDIA
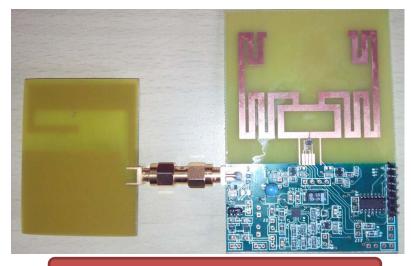
A. Robb, J. Bommer, R. Martinez, J. Harrigan, S. Ramamurthy , H. Muniganti, V. Mannangi, and KJ Vinoy, "Wireless Aircraft Fuel Quantity Indication System," 2014 IEEE Sensors Applications Symposium - , Feb 18-20, 2014, Queenstown, Newzealand.

4. RFID Integrated with Sensor

Harvesting Antenna (PIFA / Monopole)

Tag Antenna and MONZA-X2K (Meander Line)

RF-DC

Lumped Element Based Matching Circuit

Two Stage Dickson Voltage Multiplier

Power regulation and supervisory

DC-DC Converter (BQ 25504)

Power Gating MOSFET

LDO

Digital Section

Microcontroller (PIC 16F1823)

Sensors (Light, Temperature, Accelerometer)

I2C bus network and pullup resistors

Parts of Fabricated System

PIFA CONNECTED SENSING TAG

20.83 ohm discrete port

Lumped element of 181.39 value at 865/905 MHz (working frequency)

RFID Tag Antenna

Data processing Circuitry

DC-DC Boost Converter

RF-DC Conversion

Sandeep Rana 2014-15

Sandeep Rana, TV Prabhakar, KJ Vinoy, An Efficient Architecture for Battery-less Terminals for Internet of Things, Applied Computational Electromagnetic Conference, Guwahati, Dec 28-21, 2015

Characterization of Performance

Source	Power
RFID Reader	30 dbm
Circularly polarized antenna	8 dbi
Polarization loss	3 dbi
Monopole Antenna	5 dbi
PIFA antenna	1 dbi
Meander Antenna	0.4 dbi

Tag Antenna	EIRP (dbm) + Gr	Rg Expected (-10 dbm) and 50 % overall efficiency	Range Achieved
Monopole	40	7.5 mtr	7 mtr
PIFA	36	5 mtr	5.5 mtr
Meander Line	35.4	4.5 mtr	4.5 mtr
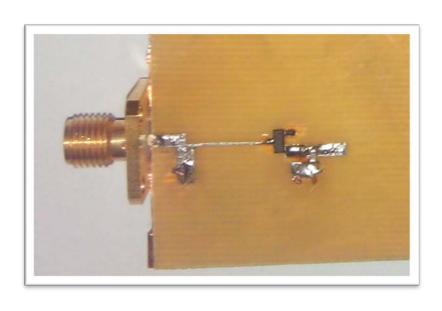
$$\lambda = 34.5 \ cm \qquad Prx = EIRP * Gr * \left(\frac{\lambda}{4*\pi*r}\right)^2$$

Efficiency worked out for -10 dbm
RF-DC efficiency – 20 % and DC-DC efficiency – 80%
Overall efficiency – 16%
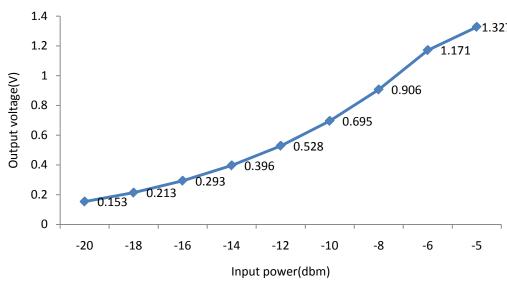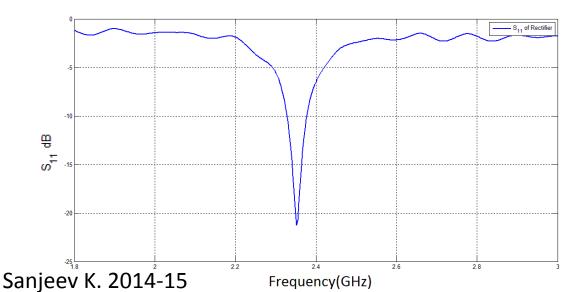
Sandeep Rana 2014-15

5. Harvesting at 2.4GHz

Output voltage vs Input power

- 0.153
- 0.213
- 0.293
- 0.396
- 0.528
- 0.695
- 0.906
- 1.171
- 1.327

Output voltage(V)

Input power(dbm)

S_{11} dB

Frequency(GHz)

S_{11} of Rectifier

Sanjeev K. 2014-15

RICOH
imagine. change.   52

Comparison of Efficiencies

Schottky diode

	900MHz	2400MHz
-15dbm	Efficiency=47% Output voltage=0.29V Load resistor=6K	Efficiency =17% Output voltage=0.2V Load resistor=6K
-20dbm	Efficiency=31% Output voltage=0.15V Load resistor=6K	Efficiency=5% Output voltage=0.1V (0.153V unloaded) Load resistor=6K
	Used HSMS 2852	Used HSMS 2862

Diode connected MOS with high Q matching

	900MHz	2400MHz
-14dbm	Efficiency=6.2% Output voltage=1.1V Load resistor=500K	Efficiency=2% Output voltage=0.632V Load resistor=500K
-20dbm	Efficiency=1.8% Output voltage=0.3V Load resistor=500K	Efficiency=0.2% Output voltage=0.118V Load resistor=500K

Zero V_{TH} CMOS

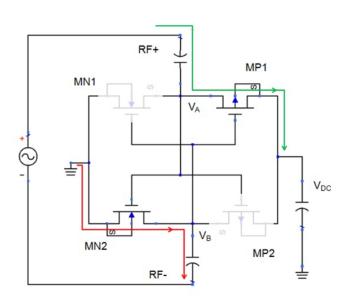|	900MHz	2400MHz
-15dbm	Efficiency=4.6% Output voltage=0.86V Load resistor=500K	Efficiency=3.79% Output voltage=0.774V Load resistor=500K
-25dbm	Efficiency=2.4% Output voltage=0.198V Load resistor=500K	Efficiency=1.7% Output voltage=0.165V Load resistor=500K

← Power

← Technologies

Frequencies →

- CMoS integration requires diodes using MoSFETs.

- Simple diode connected configurations are not effective at low power/voltage levels
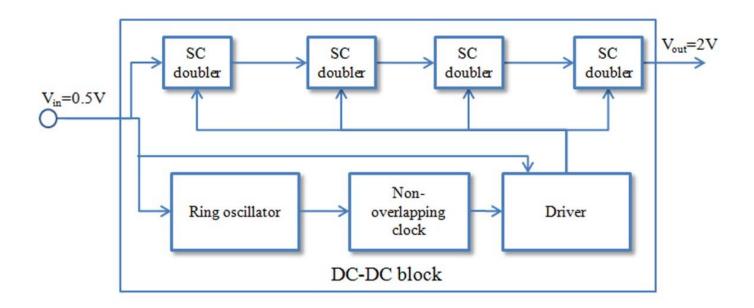
- In Cross Coupled Rectifiers

  - Biasing of MOSFETs by charge stored in capacitors. This is a way of threshold compensation.

  - Low ON resistance due to high overdrive voltage.

  - In both cycles of input, output capacitor is charged. Although DCP uses both cycles, only alternate cycles charges the output capacitor and the other cycle charges the input capacitor.

54
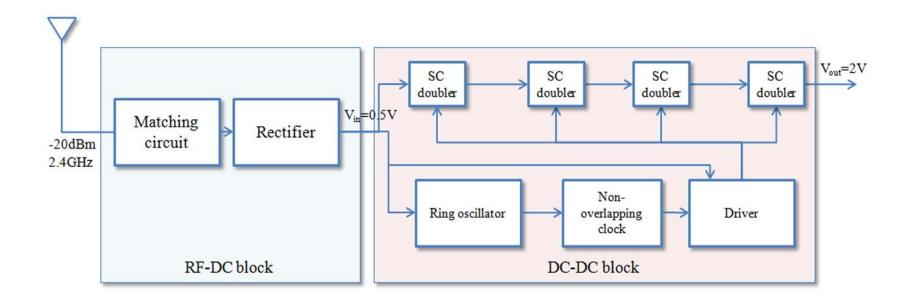
DC-DC converter

- Low loss switched capacitor DC-DC converter:

Full system block diagram

1. Output capacitor of RF-DC supplies DC-DC

2. Enable generator logic constructed using back to back inverters

3. 5 MOSFETs added to limit supply voltage to ring oscillator

Sanjeev K, Manjunath M, TV Prabhakar, KJ Vinoy, Some Practical Considerations of RF to DC Converter using Low Vth CMOS Rectifier, Applied Computational Electromagnetic Conference, Guwahati, Dec 28-21, 2015
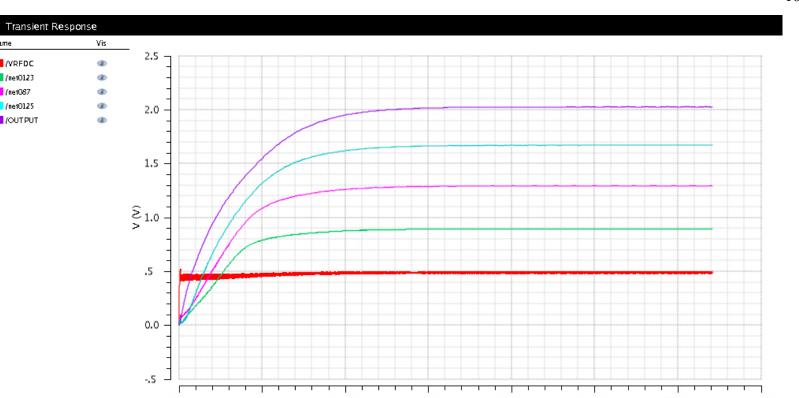
Full system simulations

- Output capacitor of RF-DC loaded heavily when clocks transition, so ripples exist in RF-DC output.

- Below 0.5V at clock transitions, above 0.5V between clock transitions

- Time step is 8ps for RF-DC simulation and DC-DC has to run for hundreds of µs or few ms, so simulations times are large.

Output voltage=2V across a load of 1.6MΩ, which gives

$$Efficiency = \frac{2^2 / 1.6 \times 10^6}{10^{-5}} = 25\%$$
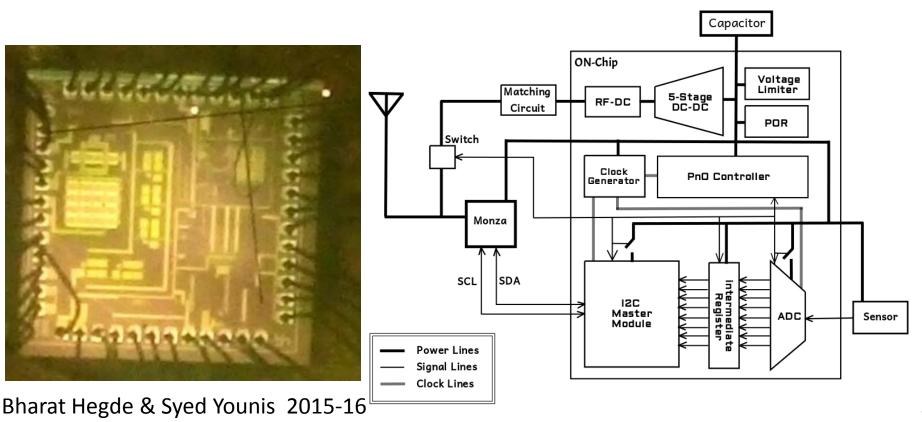
Summary with IC Design

- Low Vth NMOS based DCP gives 42.3% efficiency.

- FGCCR gives 55% efficiency.

- The overall system efficiency is 25%.

- Higher than efficiency reported in literature for RF-DC converter operating at -20dBm, 2.4GHz in 130nm technology.

Reference	This work	[5]	[7]	[6]
Power level	-20dBm	-25.7dBm	-22.6dB	-20dBm
Frequency	2.4GHz	2.45GHz	906MHz	2.4GHz
Efficiency	55%(simulated)	37%(measured)	10%(measured)	36%(simulated)
Rectifier	FGCCR in UMC 130nm CMOS	DCP in 0.5μm Silicon on Sapphire	DCP with floating gate transistors in 0.25μm CMOS	FGCCR in 130nm CMOS
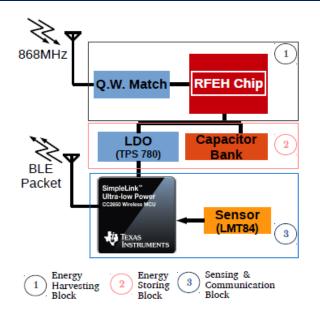
- Working on a 3-chip architecture

  - Our chip to enable sensing, and control functions

  - Communication using an external Monza chip

- Fabricated chip using commercial services!!

Bharat Hegde & Syed Younis 2015-16
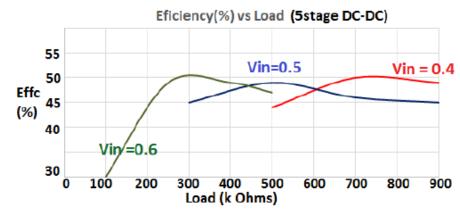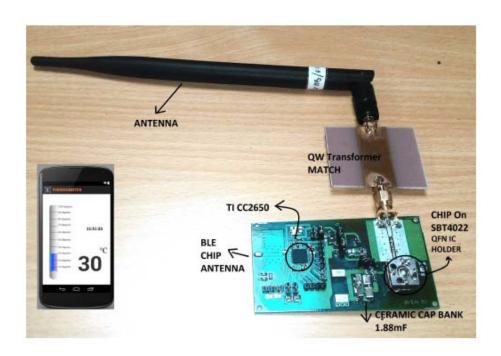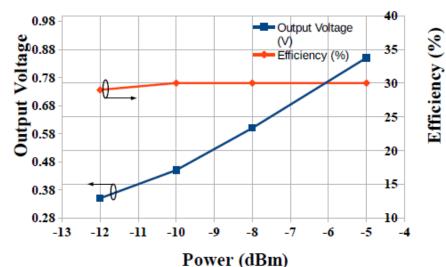
Battery-less Sensor node for BLE

Distance from Reader (m)	Input Power (dBm)	Cold Start Time (min)	Packet Time (min)	Efficiency (%)
3	-2	15	1.75	2.2
4	-5	35	7	1.7
5	-8	120	32	1

Eficiency(%) vs Load (5stage DC-DC)

Bharat Hegde & Syed Younis 2015-16

Summary

- Most low power wireless terminals operate intermittently

- These require anywhere 50uW to about 10mW for their operation.

  - Batteries limited: cost, size, stored energy

  - Solar: not dependable through

- WPT and RF EH can enable wide use of IoT

  - Main challenges in the design is the low incident energy/power/voltage

  - High Quality factor components may help

- Several fabricated examples discussed here: All can transmit data to an aggregator wirelessly

  - Different standards implemented.

Acknowledgements

- Dr. TV Prabhakar, DESE, IISc
  - Gaurav Singh
  - Rahul P
  - Chaithanya C
  - Aditya Mitra
  - Prashuk Jain
  - Prashanth Raja
  - Nirmal John
  - Nithin Jose
  - Syed Younis
  - Bharat Hegde
  - Niharika Thakuria

- Partial Funding From
  - ANRC (Boeing, Wipro, HCL)
  - Ricoh Research, India

- Prof Bharadwaj Amrutur
  - Uday S
  - NS Sreeram

- Others
  - Vivekanand M
  - Harikiran M
  - Manjunath M
  - Sanjeev K
  - Sandeep Rana

Thank YOU

Free Energy From Air Circuit

Last year I found a United States Patent that showed how to collect ambient energy right from the air. I finally decided to build this curcuit just to see what it could do.
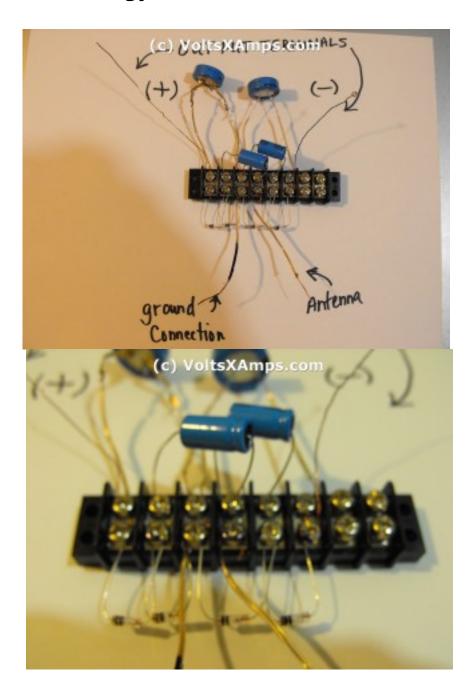
[protected]

The above images are from the circuit I built. Most of the text and info comes direct from the US4628299 filed by Joseph Tate.
 US4628299

The Amazing Ambient Power Module

Parts List for the APM-2
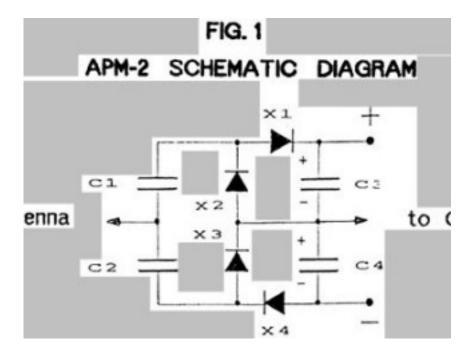
Four 1N34 germanium diodes (Radio shack #276-1123) ~ Figure 1, X1, X2, X3, & X4
 Two 0.2 mfd 50 V ceramic capacitors ~ Figure 1, C1 & C2
 Two 100 mfd 50V electrolytic capacitors (Radio Shack #272-1016) ~ Figure 1, C3 & C4
 Copper wire for antenna & ground connections

Introduction

The Ambient Power Module (APM) is a simple electronic circuit which, when connected to antenna and earth ground, will deliver low voltage up to several milliwatts. The amount of voltage and power will be determined by local radio noise levels and antenna dimensions
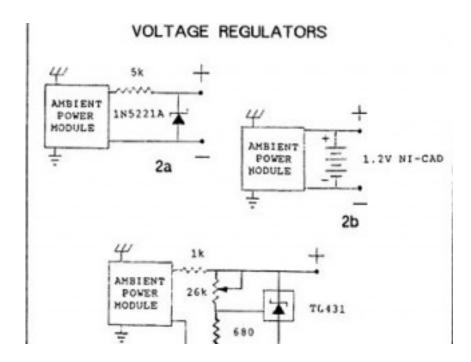
Generally a long wire antenna about 100' long and elevated in a horizontal position about 30' above ground works best. A longer antenna may be required in some locations. Any type copper wire, insulated or not, may be used for the antenna. More details about the antenna and ground will be discussed further on.

The actual circuit consists of two oppositely polarized voltage doublers (Figure 1). The DC output of each doubler is connected in series with the other to maximize voltage without using transformers. Single voltage doublers were often found in older TV sets for converting 120 VAC to 240 VDC. In the TV circuit the operating frequency is 60 Hz.

FIG. 1

APM-2 SCHEMATIC DIAGRAM

The APM operates at radio frequencies, receiving most of its power from below 1 MHz. The basic circuit may be combined with a variety of voltage regulation schemes, some of which are shown in Figure 2. Using the APM-2 to charge small NiCad batteries provides effective voltage regulation as well as convenient electrical storage. This is accomplished by connecting the APM-2 as shown in Figure 2B.

VOLTAGE REGULATORS

Charging lead acid batteries is not practical because their internal leakage is too high for the APM to keep up with. Similarly, this system will not provide enough power for incandescent lights except in areas of very high radio noise.

It can be used to power small electronic devices with CMOS circuitry, like clocks and calculators. Smoke alarms and low voltage LEDs also can be powered by the APM.

Figure 3 is a characteristic APM power curve measured using various loads from 0-19 kOhm. This unit was operating from a 100' horizontal wire about 25' high in Sausalito CA. As can be seen from the plot, power drops rapidly as the load resistance decrease from 2 kOhm. This means that low voltage, high impedance devices, like digital clocks, calculators and smoke alarms are the most likely applications for this power source. Some applications are shown in Figures 4 through 7.

FIG. 3   AMP Power Envelope
Volts vs. Resistance
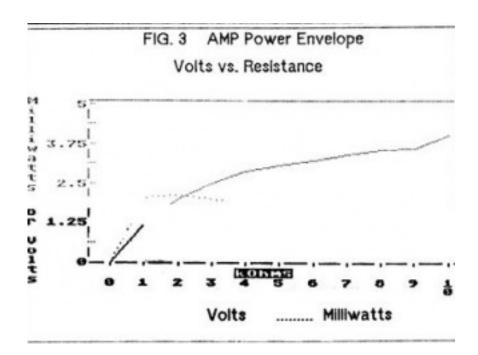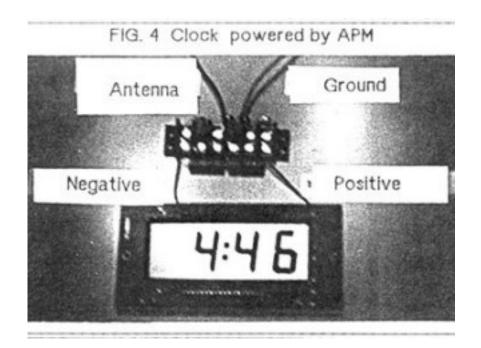
Figure 4 ~ A digital clock is shown powered by the APM-2. The 1.5 volt clock draws 28 microamps. Its position on the power envelope curve would be off the scale to the right and almost on the bottom line, dissipating only 42 microwatts.

FIG. 4 Clock powered by APM

FIG. 5 APM charging a battery
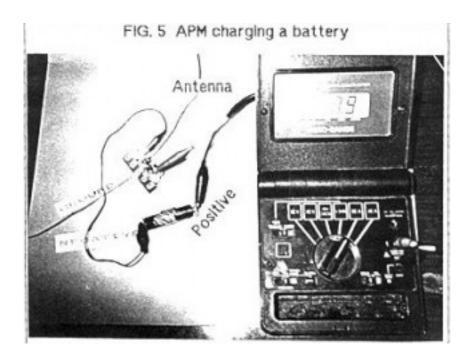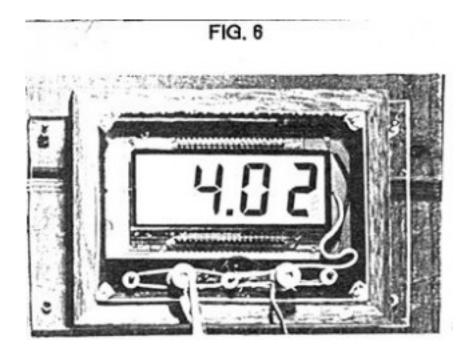
Figure 6 shows a clock which has the APM-2 built into it so it is only necessary to connect the antenna and ground wires directly to the clock. The antenna for this clock, which is a low frequency marine type, is shown in Figure 7.These antenna are expensive, not generally available, and usually don't work any better than the long wire mentioned above. But it may be necessary to use them in urban areas where space is limited and radio noise is high.
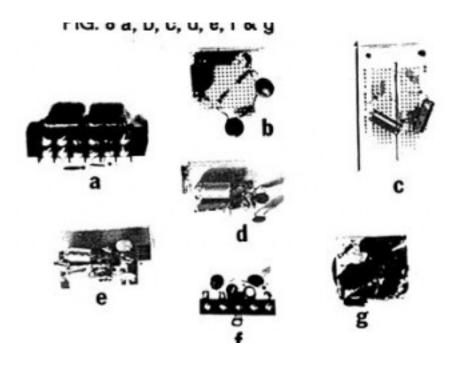
FIG. 6

Building the Module
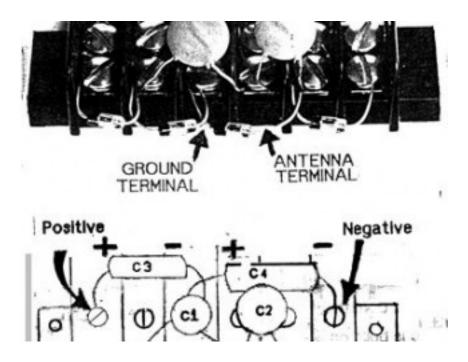
The builder has a choice of wiring techniques which may be used to construct the module. It may be hand wired onto a terminal strip, laid out on a bread board, experiment board, or printed circuit. Figure 8 shows some of the different ways of constructing the APM-2.

Figure 8A is constructed on a screw strip terminal; Figure 8B is constructed on a perforated breadboard; Figure 8C is built on a standard experiment board; Figures 8D, 8E, and 8F are all printed circuits; Figure 8F is made up on a solder strip terminal.

FIG. 8 a, b, c, d, e, f & g
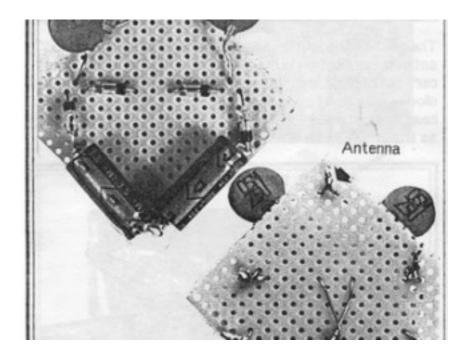
If you wish to make only one or two units, hand wiring will be most practical, either on a terminal strip or breadboard. Assembly on the terminal strip (Figure 8A) can be done easily and without soldering. It is important to get the polarity correct on the electrolytic capacitor. The arrow printed on the side of the capacitor points to negative.
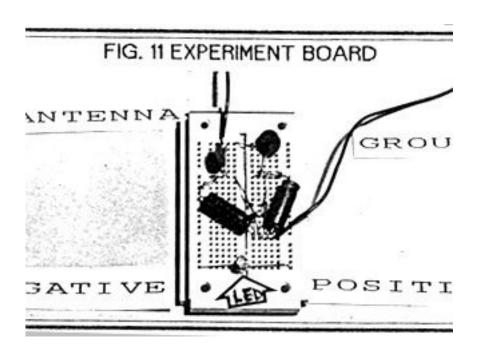
Figure 9 is a closer view of the terminal strip with an illustration of the components and how they are connected.
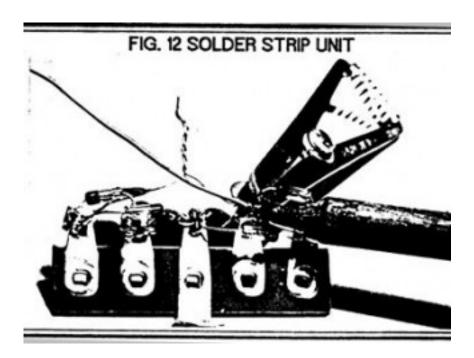
The breadboard unit is shown in Figure 10 with all components on one side and all connections on the other. All you need is a 2" x 2" piece of perforated breadboard (Radio Shack #276-1395) and the components on the parts list. Push component wires through the holes and twist them together on the other side. Just follow the pattern in the photo, making sure to observe the correct polarity on the electrolytic capacitors and the diodes. The ceramic capacitors may be inserted in either direction.

The experiment board unit is assembled by simply pushing the component leads into the board as shown in Figure 11. This unit is powering a small red LED indicated by the arrow.

FIG. 11 EXPERIMENT BOARD

The solder strip unit is made up on a five terminal strip. The antenna connection is made to the twisted ends of the ceramic capacitors. When soldering the leads of the 1N34 diodes, care must be taken to avoid overheating. Clip a heat sink onto the lead between the diode and the terminal as shown in Figure 12.

FIG. 12 SOLDER STRIP UNIT

FIG. 13
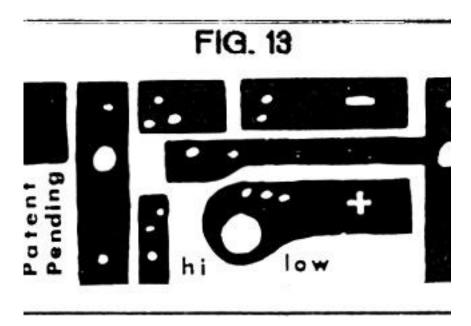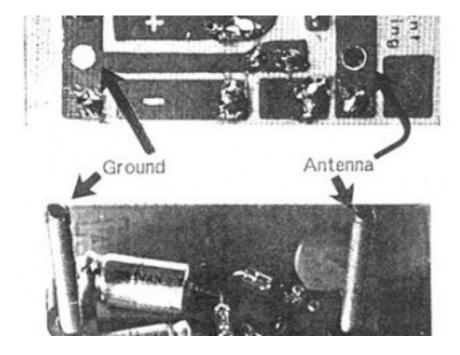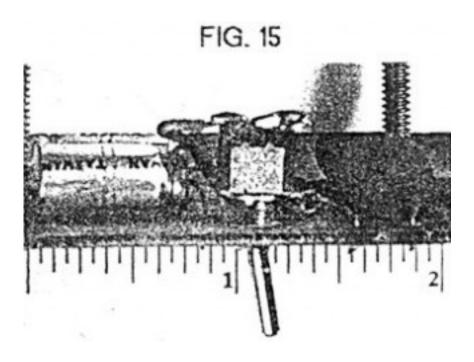
Figure 14 shows the front and back view of the completed printed circuit.

A small switch may be installed on the board to activate the zener regulator (Figure 15). This board was designed for use in clocks.

FIG. 15

Antenna Requirements

The antenna needs to be of sufficient size to supply the APM with enough RF current to cause conduction in the germanium diodes and charge the ground coupling capacitors. It has been found that a long horizontal wire works best. It will work better when raised higher. Usually 20-30 feet is required. Lower elevations will work, but a longer wire may be necessary.

In most location, possible supporting structures already exist. The wire may be stretched between the top of a building and some nearby tree or telephone pole. If live wires are present on the building or pole, care should be taken to keep your antenna and body well clear of these hazards.

To mount the wire, standard commercial insulators may be sued as well as homemade devices. Plastic pipe makes an excellent antenna insulator. Synthetic rope also works very well, and has the advantage of being secured simply by tying a knot. It is convenient to mount a pulley at some elevated point so the antenna wire may be pulled up to it using the rope which doubles as an insulator (Figure 16).
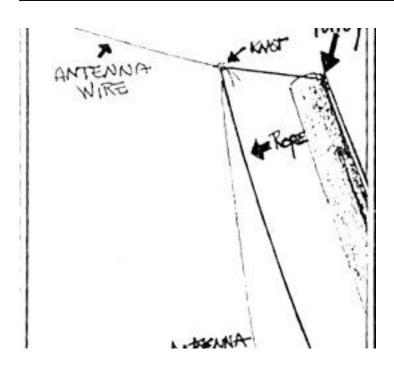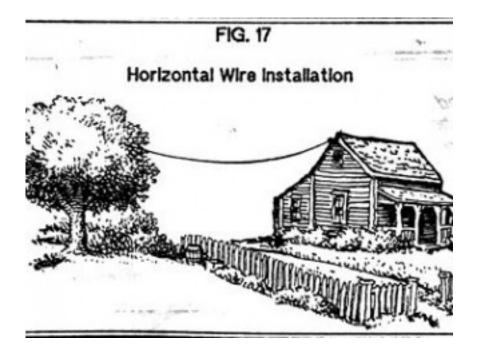
Figure 17 is an illustration of a horizontal wire antenna using a building and tree for supports.

Grounding

Usually a good ground can be established by connecting a wire to the water or gas pipes of a building. Solder or screw the

wire to the APM-2 ground terminal. In buildings with plastic pipes or joints, some other hookup must be used. A metal rod or pipe may be driven into the ground in a shady location where the earth usually is damper. Special copper coated steel rods are made for grounds which have the advantage of good bonding to copper wire. A ground of this type usually is found within the electrical system of most buildings.

Conduit is a convenient ground provided that the conduit is properly grounded. This may be checked with an ohmmeter by testing continuity between the conduit and system ground (ground rod). Just as with the antenna, keep the ground wire away form the hot wires. The APM's ground wire may pass through conduit with other wires but should only be installed by qualified personnel.

Grounding in extremely dry ground can be enhanced by burying some salts around the rod. The slats will increase the conductivity of the ground and also help retain water. More information on this subject may be found in an antenna handbook.

Good luck getting your Ambient Power Module working. It is our hope that experimenters will find new applications and improve the power capabilities of the APM.
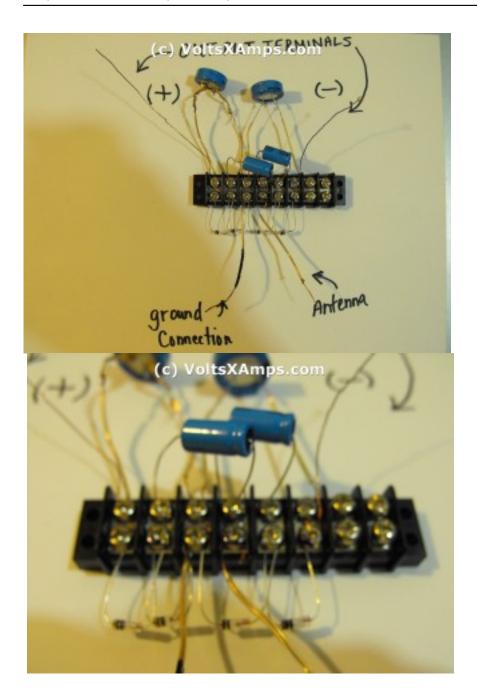
CopyRight Joseph Tate
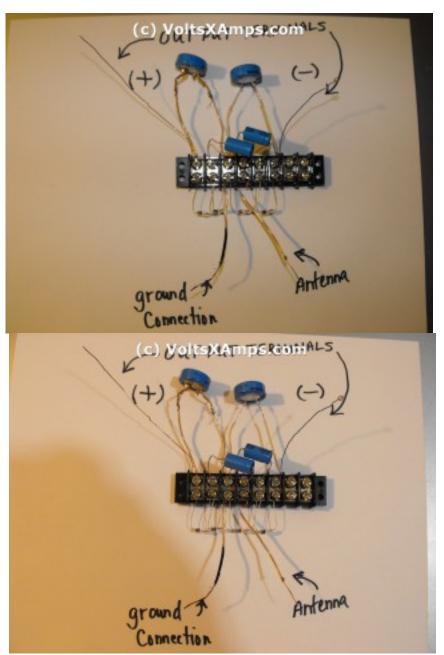
 Inventors: Tate, Joseph B. (Sausalito, CA)
 Brown, David E. (Mill Valley, CA)

 Application Number: 06/695632
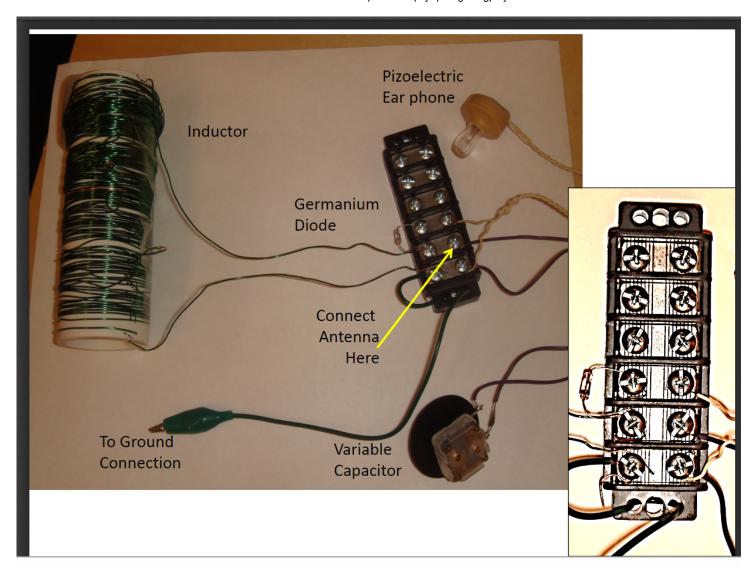
 Publication Date: 12/09/1986

 Filing Date: 01/28/1985 NOTE: The images maybe hard to see as they were originally scanned and uploaded in black and white. If it helps you to replicate this device you may want to check the images of the one I built at the very top (in color) Here are some more photos of my completed unit for your review.

[/protected]

Inductor

Pizoelectric
Ear phone

Germanium
Diode

Connect
Antenna
Here

To Ground
Connection

Variable
Capacitor
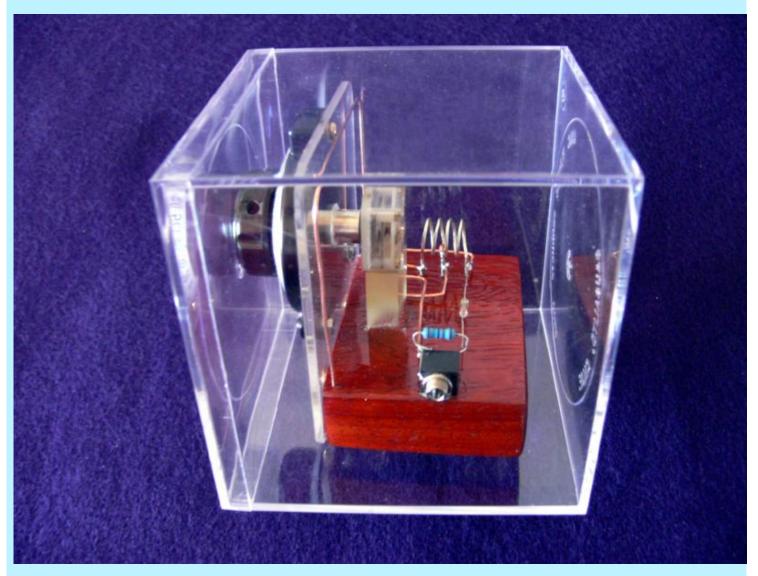
FM Crystal Radios?

copyright © 2006 by Larry J Solomon

I have heard, even from a physicist, that it is impossible to build FM crystal radios. On the other hand some experimenters claim that they have built them. This argument intrigued me to try and build an FM crystal radio, which I have done successfully. To my surprise, the result is an astounding performer, pulling in four local stations in Tucson. When connected as a receiver to a good sound system the sound fidelity is as good or better than more expensive AM radios. In fact, it sounds "high-fidelity".

This picture shows the Solomon FM Crystal Set in an acrylic display case. I made the set specifically to fit inside this case (the case came first).

My definition of a crystal radio is one that is not powered, except by the radio transmission itself and employs a crystal detector. So, it should work without any batteries or AC power. An FM crystal receiver must be able to detect and receive FM signals well enough to be heard in earphones without any such extra power.

This FM receiver is an amazing performer. It has crystal clear reception (pun intended), good sensitivity, but only fair selectivity. This set was a discovery for me. I started out by designing and building the normal AM sets. Then one day while testing the "Mystery" set (see my other web links), to my surprise, in addition to the expected panoply of AM stations, I heard a very faint signal that I could not tune out. At first, it seemed too weak to identify. When I tuned out all the AM stations, I was astonished to hear the announcement "KiiM FM, 99.5"! This is a

country music FM station here in Tucson. It was all over the dial, untunable, but the much louder AM signals masked it when they were tuned in.
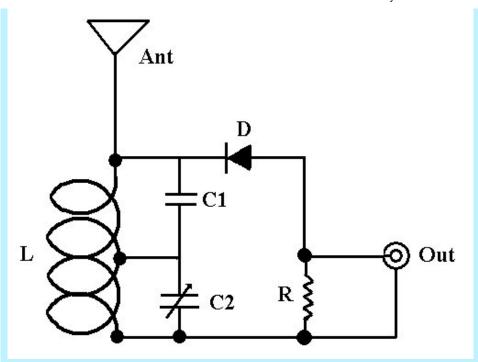
I set myself the task of trying to improve the FM reception. I tried some simple circuit modifications that did not seem to improve anything. Then I connected a dipole antenna instead of the AM antenna I normally use. Suddenly, the FM signal was much clearer, although still weak. By using the audio output and sound system amplifier, I was even more amazed that four different FM stations came in loud (or rather medium) and clear. I found that changing the telescoping antenna length and position I could tune the stations in and out. They were KRQ, KLPX, KiiM, and KHYT all local FM stations with transmitters nearby. Their reception was also affected by the length and position of the audio output cable.

After doing some research, I discovered that there was a physical theory that claimed that FM reception was possible and even probable using the same circuit as an AM receiver. The theory is called "slope detection". So, I set out to find circuit improvements. A web search yielded little, mostly theory. But there was enough information that I thought I could make some modifications to the AM circuits to make them more tunable to FM signals and less tunable to AM. Since FM operates at higher frequencies, all I had to do, I thought, was make the coil and caps smaller. After much "tinkering" I arrived at the current circuit.

The circuit looks identical to a classic AM crystal circuit but is even simpler to build. The components were reduced in dimension to resonate at higher frequencies. This was done by experimenting with smaller and smaller coils and capacitors. The antenna is also much reduced in size (from that of AM) to resonate at higher frequencies (the antenna is crucial). The air variable capacitor I used has two trimmers in it which should be adjusted for best reception.I have found that a commonly available vernier dial and knob will fit the capacitor nicely. See end of article for a picture of the variable. C3 is a ceramic capacitor of 18 pf, but may be anywhere from 10 to 50pf. A detected FM signal is converted to AM due to an effect called slope detection that modulates amplitude.

This FM Crystal Set works best near the transmitter (I have not tested it beyond about 10 miles). Secondly, the sound level is quiet, especially without an amplifier. A quiet room is needed for listening with earphones. One must be willing to move the set around to find a location for the best reception of signals. However, in addition to listening with high impedance earphones (crystal or otherwise), the set can be connected directly to an audio amplifier's low level magnetic input which can then play amplified through a sound system at any volume -- sounds GREAT. In fact, I recommend starting tests with the FM crystal set by connecting it to the low-level phono inputs of a receiver or preamplifier. (Nowadays, many receivers don't even have a phono input!) That way you can crank up the volume, which makes it more likely to find the FM stations. If no signals are detected, I also recommend connecting an external "rabbit ear" antenna or hanging a short wire (12 inches or so) in various positions next to the internal antenna. The variable length of rabbitt ears can help to tune in stations.

No additional wiring or antenna is necessary (the antenna is optimized in length for FM.)

L - 4 turns #18 copper or silver wire, 12mm inside diameter, tapped at 2.5 turns
Ant - 7 inches of #18 bare copper wire
C1 - 18 pf ceramic capacitor
C2 - 50 pf air variable capacitor
D - 1N34 diode or rock crystal
R - 150K resistor
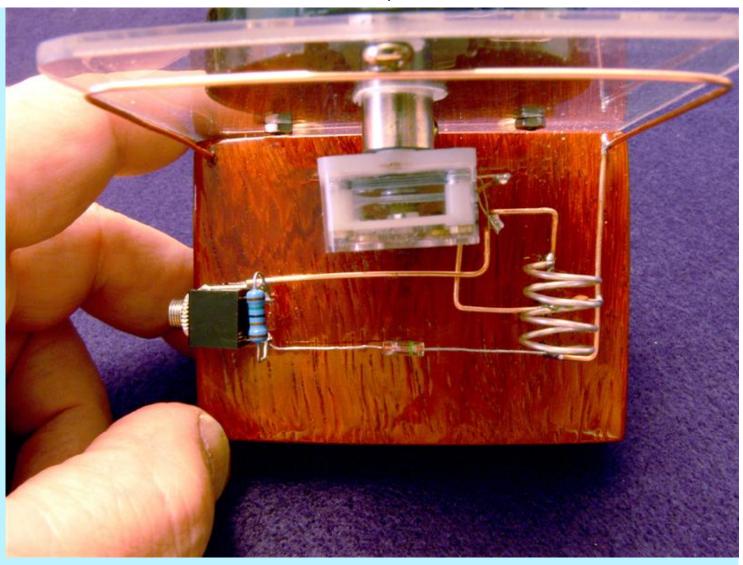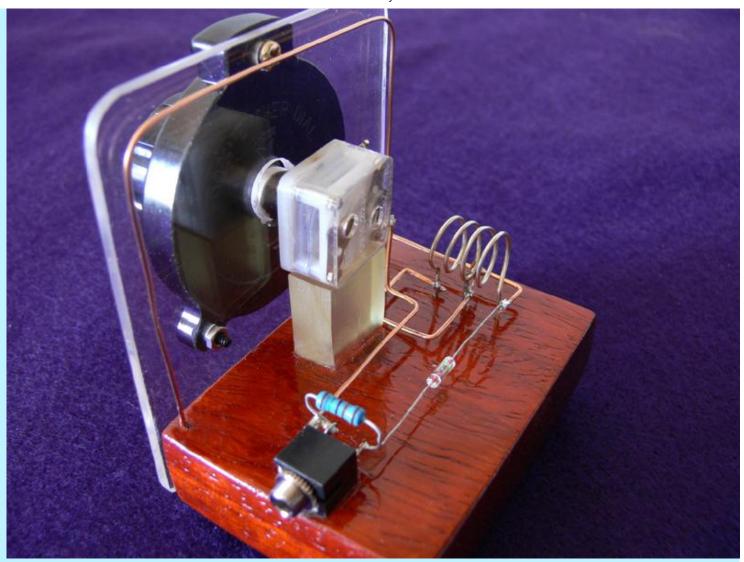
The diode is tapped directly to the antenna. The vernier dial fits directly on the tuning capacitor. The antenna parallels the perimeter of the acrylic face plate. "Military style" #18 AWG wiring is used without any insulation. It is important to keep the components physically close together. The component specifications are the same as in circuit #2. The coil is silver rather than copper, but copper does just as well. I think that the contrast of the silver and copper is beautiful. The coil was wrapped around a Sharpie Permanent Marker, then slipped off and expanded slightly. The wooden base is made from lacquered, polyurethane padouk.

I consider this set a work of art as well as science and think it is the most elegant crystal receiver I have created. I love the contrast of the silver coil, the copper antenna, the clear acrylic faceplate, the black vernier dial, the white and transparent variable capacitor, and the subtle colorings on the resistor, the diode, and the lucite base. Yet the circuit is so ridiculously simple that some will not believe it is possible without building it themselves. No shielding is necessary, and there is no problem with hand capacitance. However, the output cable position may affect reception sensitivity.
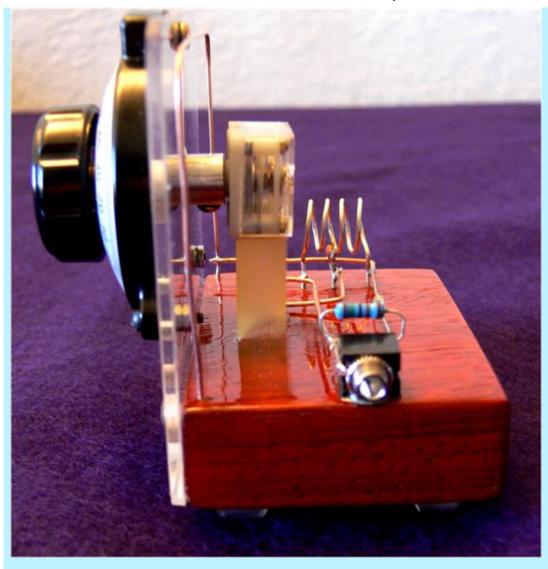
Photos of wired circuit

A hand is included in this photograph to show scale. Note the military style wiring, diode, and antenna. I wanted the wiring to create a modern design similar to a Mondrian painting. Not only is this set beautiful, it works! No power and no long antenna! It looks like a work of fiction.

Is this thing imaginary -- science fiction? Well, imagination did play a part, but it is definitely not science fiction. This shot shows the elegance of the FM set best, I think. There is only one resistor and one fixed capacitor.

The inside of the tuning capacitor and the phono jack/output can be seen here. Can you spot the fixed ceramic capacitor? Note the polished edge of the face plate and the reflection in the wooden base.
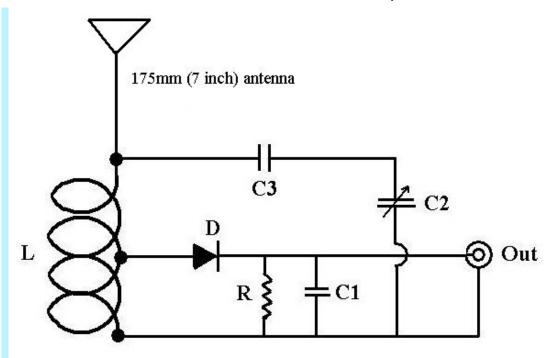
A quarter-inch piece of lucite was fitted under the tuning capacitor to anchor it. Note the two tiny trimmers on the back of the tuning capacitor. Brass screws were used to enhance appearance.

The vernier dial is large to accomodate ease of tuning, and the vernier makes it easy to separate stations. Two golden (brass) wood screws fix the face plate to the base. Holes for the face plate were made with special plastic drills, but ordinary drills may be used if drilled very SLOWLY. The knob is removable.

FM Crystal Circuit #2

175mm (7 inch) antenna

L - 5 turns AWG#18 bare copper or silver wire, 12mm inside diameter, tapped at 2.5 turns
D - 1N34 or rock crystal diode
C1 - 82 pf capacitor
C2 - 80 pf air variable capacitor
C3 - 18 pf capacitor
R - 150K resistor
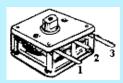
The following photographs show the circuit wired with the handmade Saturn Dial. and knob. It is perhaps not as visually striking as set No. 1, but it works just as well. In fact, this set was the original version. Notice that all the wiring and coil are copper.

The Saturn dial and knob were fashioned from a "doll's head" from Michael's Arts and Crafts, a piece of lucite cut with two circle cutters, and a brass paper fastener. The knob is fixed to the tuning capacitor with a small machine screw that fits in the hole below the brass fastener. The most difficult part of this was fashioning "Saturn's rings". This must be done very carefully and slowly. The inside edge should be cut slightly undersized and then sanded with a drum sander to fit snugly. The outside edges can be sanded with fine sandpaper and polished with a plastic polisher.

The air variable capacitor may be obtained from Electronix Express at http://www.elexp.com/. Part number 14VCRF10-280P. The 80 pf side is recommended for the second circuit, contacts 2-3. Contacts 1 and 3 were used for the first circuit (50pf).

- OSC: 5-59 pf
- ANT: 5-142 pf
- OSC and ANT Trimmer 10pf range

Home -- If you want to contact me go to the home page and go through the Music Clinic link.

AUTODESK. Make anything. (http://www.autodesk.com)

FREE ENERGY FROM THIN AIR!

Technology (/technology/) > Science | by DrewPaulDesigns (/member/DrewPaulDesigns/) Follow
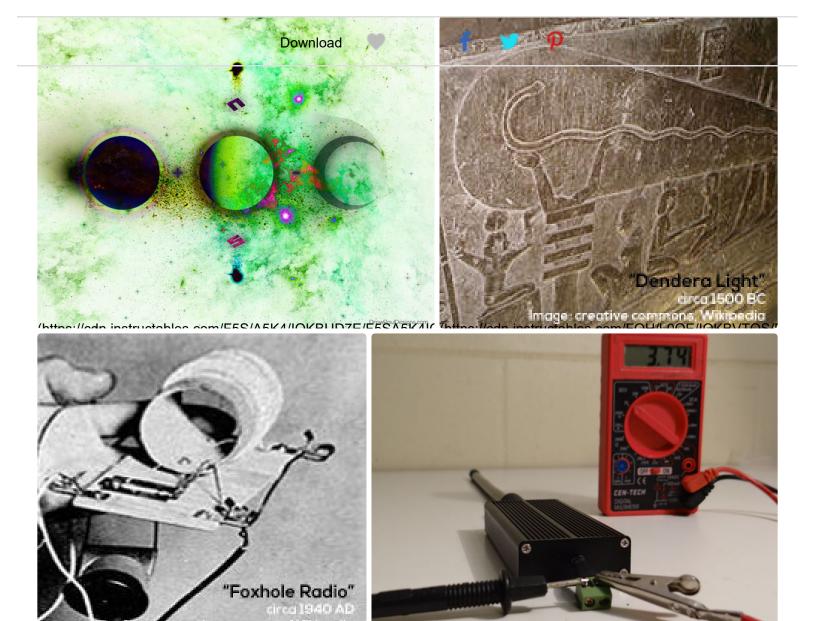
79,181 1,078 230

Posted May. 24, 2016 | (cc) BY-NC-SA

FREE ENERGY FROM THIN AIR!

*INVOLVES CRYSTALS AND DANGEROUS VOLTAGES!

"Dendera Light"
circa 1500 BC
Image: creative commons, Wikipedia

"Foxhole Radio"
circa 1940 AD
Image: creative commons, Wikipedia

"Free energy from the air?, Yea, right!" Sardonic skepticism was my first reaction to this unusual concept, as well.

Though, its not so far out there, in fact. Light can be converted to DC current with solar panels, electricity can be converted to magnetism as I did in my last article (https://www.instructables.com/id/DIY-Electro-Magnetic-Levitation/), in a microphone sound waves are converted to an electrical signal (by vibrating a magnet near a coil (http://hyperphysics.phy-astr.gsu.edu/hbase/audio/mic.html)), solar rays can even be focused and converted to heat in awesome devices like this! (http://www.gosunstove.com/) When we think about it, energy is all around us and can be harvested in an enumerable many of ways.

Today, we are going to take a rather novel approach. We are going to build a device specifically designed to sense and capture a particular band of energy which is all around us.

The earth is magnetic and anyone who has ever used a compass knows this. Magnetic bodies in motion produce electricity, we can see this in any alternator, like the one in your car. So, therefore the earth is electric as well as magnetic, by definition.

Can we detect this energy? Yes, we sure can! Ever turn on a radio in the middle of nowhere and heard static? That is your radio picking up naturally occurring energy in the RF spectrum!

Can we use this energy to do work? Absolutely! This has been known for a long time. Crystal radios (https://en.wikipedia.org/wiki/Crystal_radio) have been around since before the 1930's and can run with no input energy other than the radio signal. Even when completely isolated, but from the atmosphere, a crystal radio will produce a voltage in the earpiece resulting in a sound (albeit and undesirable one).

Well, this is where it gets interesting...

Can we replicate this effect? Yea, and with modern components like the high quality crystals found in germanium diodes, we can even increase efficiency. By applying this concept as a Crystal Energy Receiver we can take advantage of a wide range of energetic frequencies rather than tuning in to just one.

Can we scale it up? Definitely. Things like micro germanium diodes, high efficiency antennas and compact contemporary capacitors make the components that are required to build a crystal receiver fit in the palm of your hand. While there may or may not be a more efficient way, this renewable energy solution is simple to employ

and can be scaled up, or down, indefinitely.

f y p
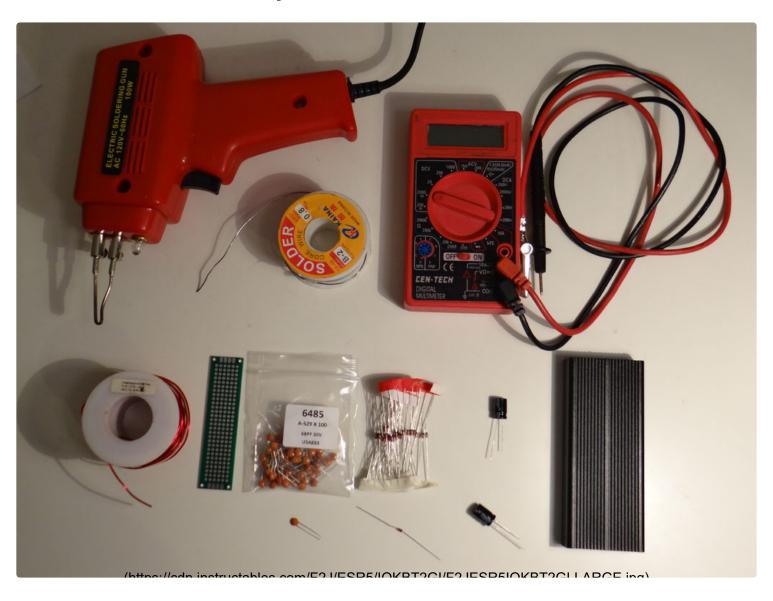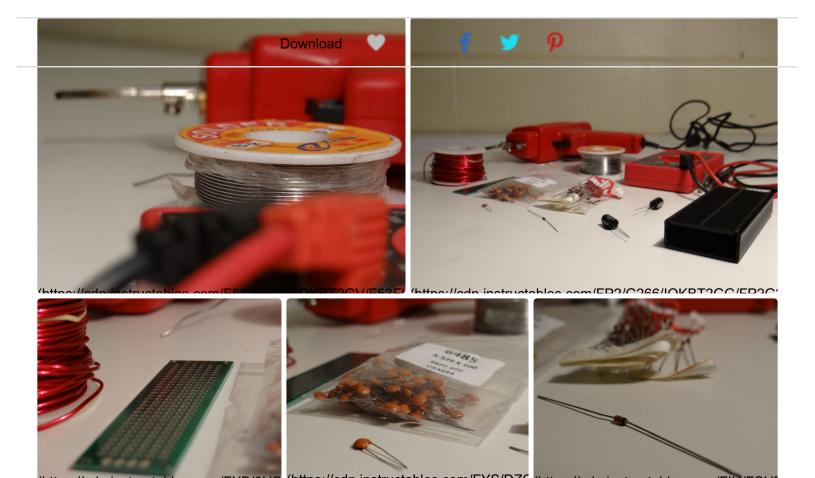
It sounds like we can build a Crystal Energy Receiver. Let's give it a shot...
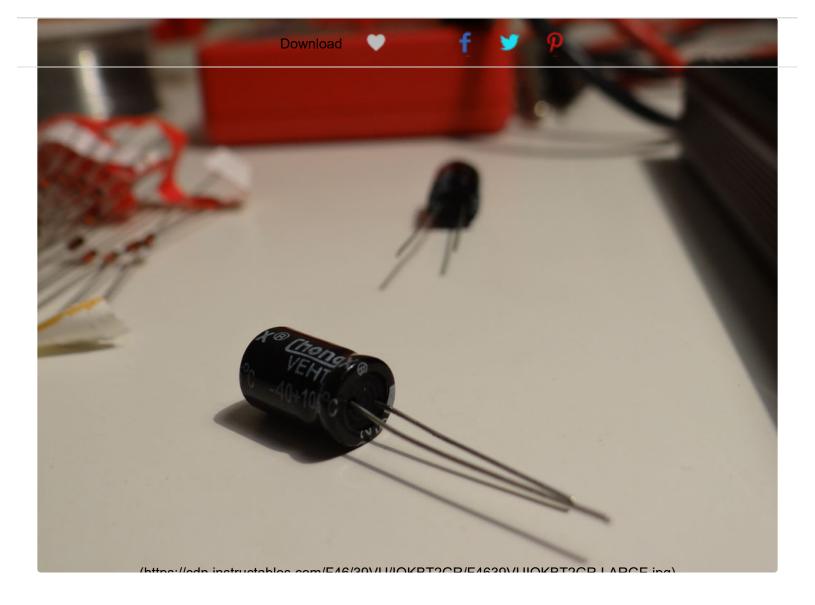
Add Tip   Ask Question
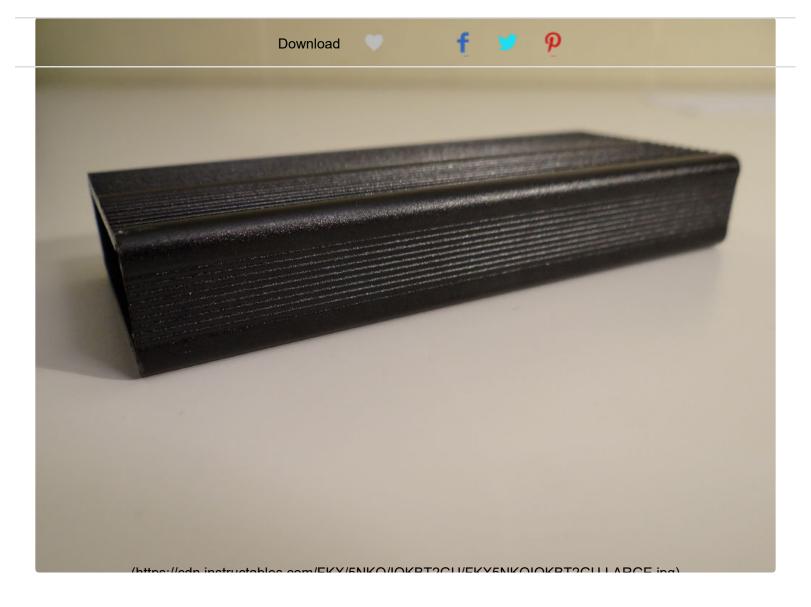
Step 1: What You'll Need

(https://cdn.instructables.com/F2.I/FSB5/JQKBT2GI/F2.IFSB5JQKBT2GI.LARGE.jpg)

(https://cdn.instructables.com/FKY/5NKO/IOKBT2CU/FKY5NKOIOKBT2CU.LARGE.jpg)

One of the reasons this particular renewable energy harvesting method is so viable is the relatively few and easy to obtain materials required.

The simplest crystal receiver design needs no power and can be built with only three parts: a coil, a crystal and a resistor. We're going to optimize that design in order to produce a cleaner and more reliable output signal by first polarizing the input amplitude, then rectifying and filtering the signal. Then we'll add an antenna, case and connections.

Get the circuit diagram here (http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html)
Get the kit here (http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html)
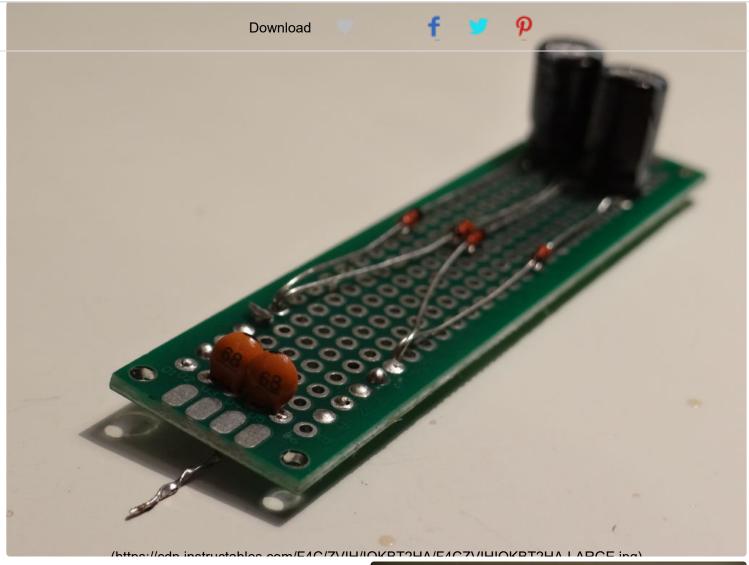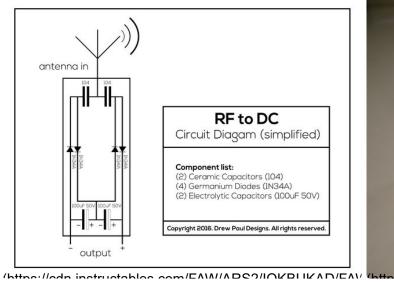
The parts for the circuit include:

(1) Circuit Board (http://www.drewpauldesigns.com/crystal-energy-receiver-kit.html)

(1) 10-18 gauge Copper Wire

(2-12+) Ceramic Capacitors (matched)

(2-6+) Electrolytic Capacitors (matched)

*note various types of capacitors can be used

(4) Germanium Crystal Diodes (1A+)

Total Unit Cost: +/- $0.40 (USD, scaled for volume of 1,000+ units)

In addition, you'll probably want to get:

(1) Project box (optional)

(1) Antenna (a loop antenna or elevated antenna is recommended and can be made with copper wire)

The tools you'll need are:

Soldering Iron/ Solder (optional)

Multimeter

Oscilloscope (http://www.seeedstudio.com/depot/DSO-Nano-v3-p-1358.html?cPath=63_65) (optional)
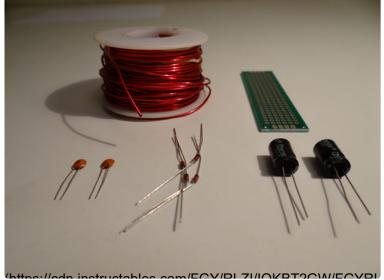
That's it. Yup, that's all. Once we've got it all, let's begin.
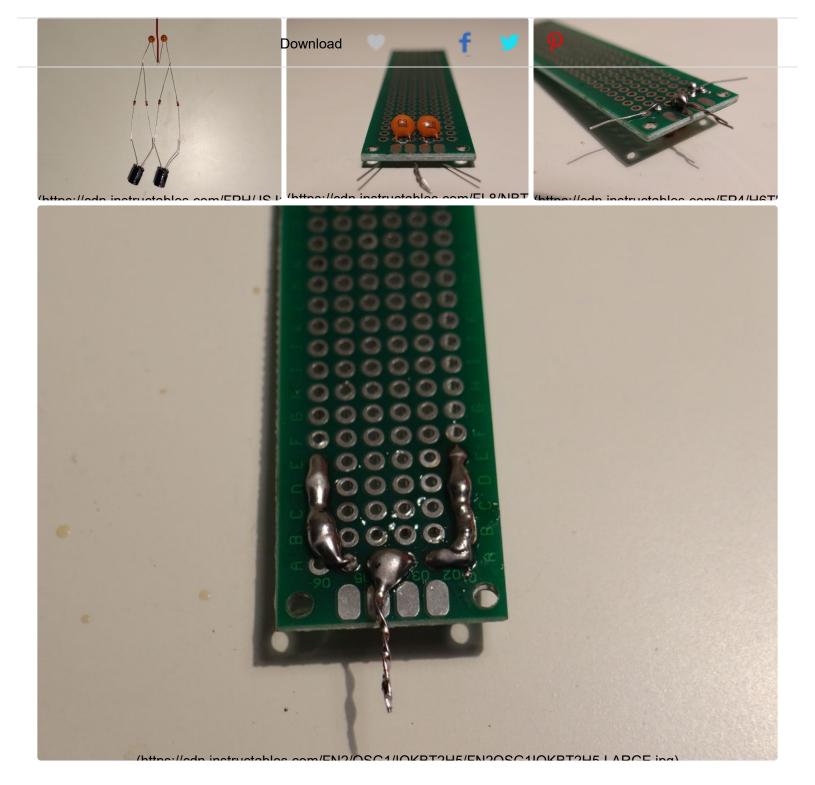
Add Tip   Ask Question

Step 2: Build the Circuit

(https://cdn.instructables.com/F4C/ZVIH/IOKBT2HA/F4CZVIHIOKBT2HA.LARGE.jpg)

antenna in

104   104

RF to DC
Circuit Diagam (simplified)

Component list:
(2) Ceramic Capacitors (104)
(4) Germanium Diodes (1N34A)
(2) Electrolytic Capacitors (100uF 50V)

Copyright 2016. Drew Paul Designs. All rights reserved.

1N34A   1N34A   1N34A   1N34A

100uF 50V 100uF 50V

−   + −   +

output
−   +

(https://cdn.instructables.com/F4W/ABS2/IOKBUKAD/F4V   (https://cdn.instructables.com/FGX/RLZV/IOKBT2GW/FGXRL

(https://cdn.instructables.com/FRH/JS...

(https://cdn.instructables.com/FL8/NRT...

(https://cdn.instructables.com/FR4/H6T...

(https://cdn.instructables.com/FN2/OSG1/JQKBT2H5/FN2OSG1JQKBT2H5.LARGE.jpg)

We're going to build the simplest version of this circuit variation in order to understand how each component interacts and as a proof of concept.

There are three simple systems at work in the circuit that are composed of capacitors, which store energy, and diodes that direct it.

Energy in the band of radio waves, among others, will vibrate a wire antenna on an atomic level, sending a discernible signal to its lead. This signal will then meet the junction between two ceramic capacitors wired in series. This junction will force positive charge from the wave to travel in one direction and negative charge in the other direction which, when collected again, makes the signal uniform and polar. Connecting the two capacitors in series creates leads on each end; the now positively charged side of one and the now negatively charged side of the other

creates a two cell battery.

This next stage of the circuit takes a signal with a net value of zero, adds the absolute values of the positive and negative amplitudes with respect to the origin and produces a positive integer. This concept can be thought of as taking:

(+1) + (-1) + (+1) + (-1) = 0

and converting it to:

| [(+1) + (+1)] | = 2
+
| [(-1) + (-1)] | = 2
= 4

Isn't math fun?

To each of these leads from our two capacitors in series, we will connect two crystal diodes, one facing each direction, to form what is called a bridge rectifier. A bridge rectifier is a configuration which will convert an alternating current to a direct one by cleverly rerouting the signal.

By connecting the bridge rectifier as shown in the circuit diagram, this direct current from the diodes then charges the electrolytic capacitors. This stage normalizes the amplitude, making the current constant and usable.
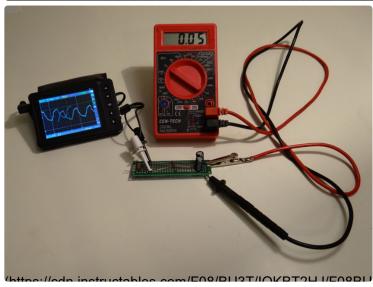
Components can easily be twisted together for testing and then soldered to a circuit board to secure.
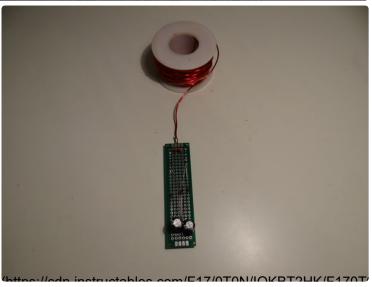
Add Tip   Ask Question
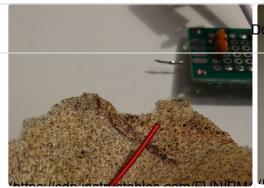
Step 3: Test and Optimize Your Circuit
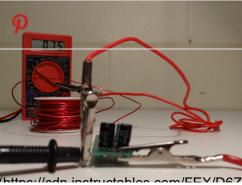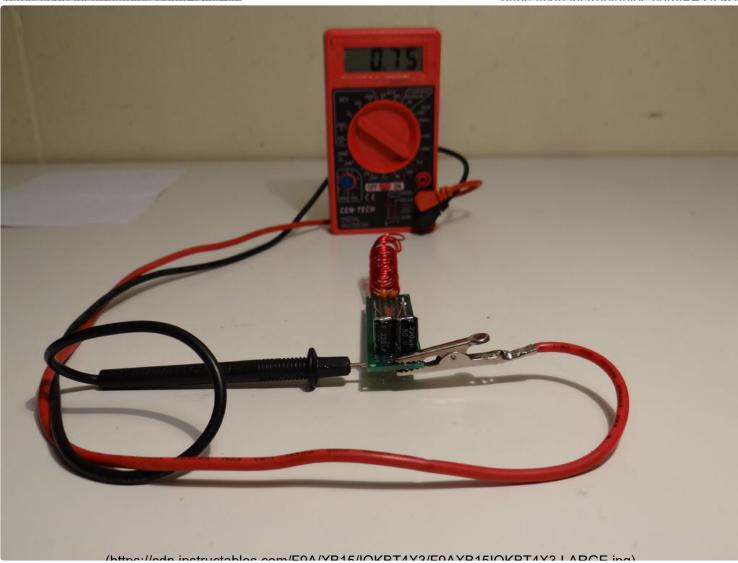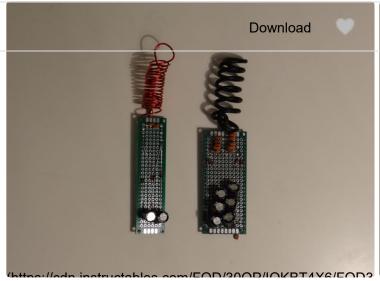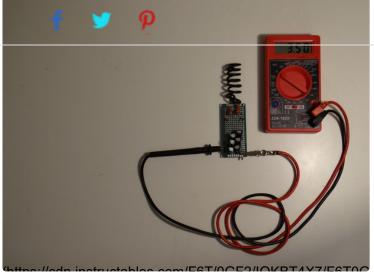
To test and analyze our circuit, we'll be using a digital voltmeter and oscilloscope.

By connecting a voltmeter to the output, we'll immediately begin to see a small voltage climbing in the 10-100mV range. If not, we'll want to check our connections and make sure the circuit is not isolated from the environment by taking it outside to a clear area.

Then, by connecting an oscilloscope to the outside leads of our two ceramic capacitor bank, we will see the the polarized signal being captured from the air around us. We can then connect after the diodes to see our varying direct current and then to after the electrolytic capacitors to see a normalized, usable direct current at our output.

We can then optimize the input resistance in two ways. Firstly, we can add additional ceramic capacitors in parallel to our original two and make sure our soldered connections are consistent and thick in this area.
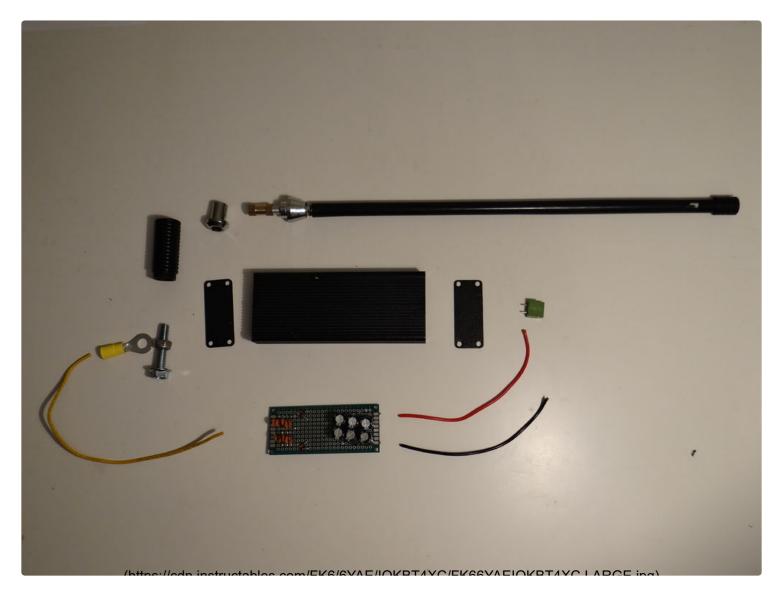
We can optimize the circuit's capacity by adding electrolytic capacitors in parallel to our original two which will allow this circuit to charge slightly when not in use. For this purpose, a charging circuit can also be added here in order to incorporate an optional battery bank.
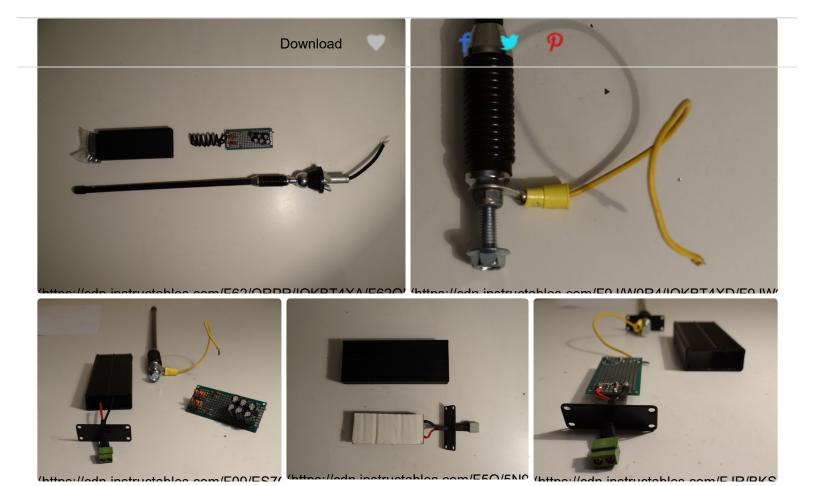
We can optimize the antenna by attaching loops and coils of copper wire in various positions, store-bought antennas or by stringing some wire up to the highest point you can reach.
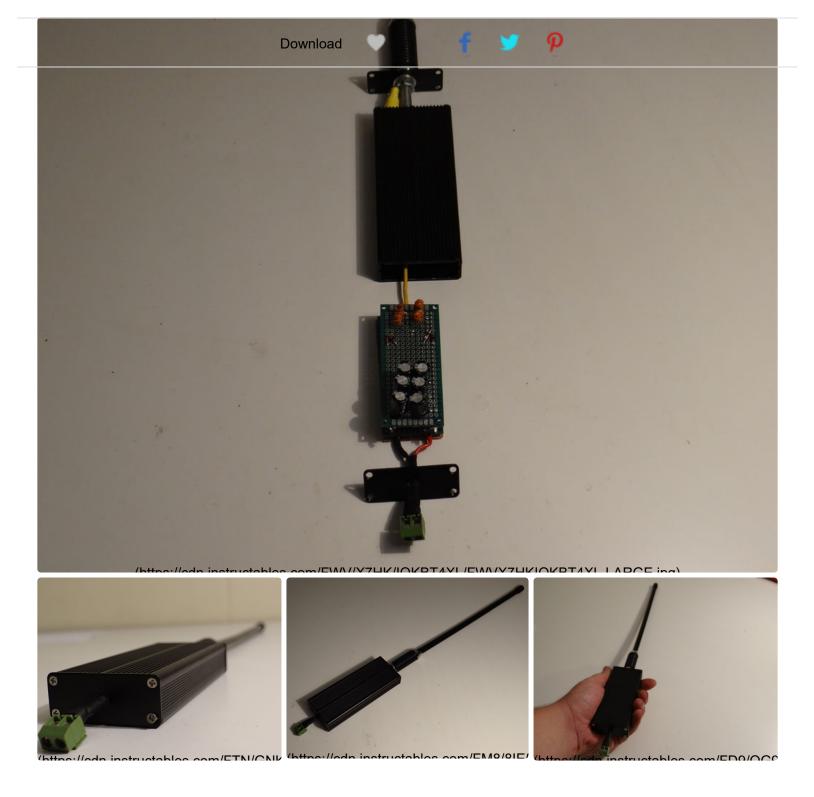
We don't have to stop there, either. We can now connect multiple circuits in series to increase voltage or in parallel to increase current. This can be done indefinitely.

Add Tip   Ask Question

Step 4: Add a Case and Antenna

(https://cdn.instructables.com/FWV/X7HK/IOKRT4YL/FWV/X7HKIOKRT4YL.LARGE.jpg)

(https://cdn.instructables.com/FTN/GNK

(https://cdn.instructables.com/FM8/8JF/

(https://cdn.instructables.com/FD9/OCS

After choosing an antenna in the last step we'll now want to permanently wire it. Whether you choose a compact antenna for portability or a tall fixed antenna for power and range, we will wire it in the same manner according to the diagram in the previous step. Note that the input on the configuration here is grounded to the metallic case, and thus the users hand, and incorporation of longer antennas will require proportionally more substantial grounding.
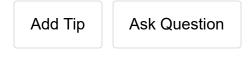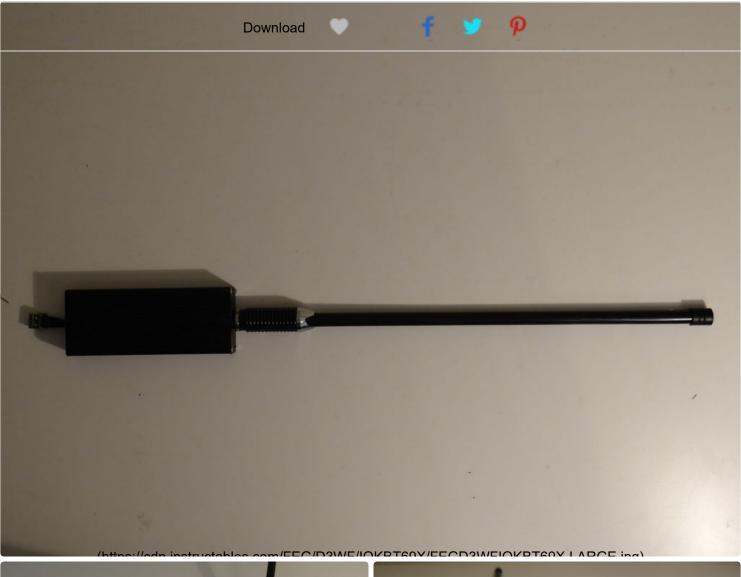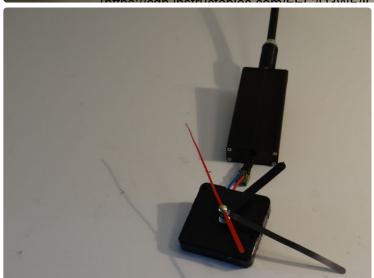
~~We will then attach a terminal to the output to allow us to connect this circuit to an~~ electrical device or charging circuit and battery bank.

Next, we will add a case, making sure to isolate exposed leads with non-conductive material especially if mounting in a conductive case. A piece of cardboard secured with glue is sufficient for the circuit's bottom and shrink wrap or electrical tape can be used in the case of any additional exposed leads. Drill two holes in your enclosure, one for the antenna or antenna lead and another for your output terminals. You can then insert your components, fasten the enclosure and your device is ready to use!

Add Tip   Ask Question

Step 5: Your Crystal Energy Receiver Is Complete!

(https://cdn.instructables.com/FFC/D3WF/IOKBT69X/FFCD3WFIOKBT69X.LARGE.jpg)

(https://cdn.instructables.com/FX4/8V5I/IOKBT4X6/FX48V5 (https://cdn.instructables.com/FLD/DTVI/IOKBT4XG/FLDDT

Your Crystal Energy Receiver is now complete and ready to use!

I built a portable version, for proof of concept and demonstration purposes. However, you can go as big as you want- to passively charge batteries or run equipment remotely; or go as small as you want- to power sensors, RFID devices, small electronics and more.

I used this harvested energy to easily power a low-consumption quartz clock, a digital chronograph with integrated circuits and LCD and was even able to momentarily rotate a small dc motor.

Because of its simplicity this device is a durable, efficient and reasonably effective method of harvesting radiant energy in a simple, replicable and sustainable way. I humbly hope that the contributions made here, and by those reading, can be one day used by people worldwide to conveniently capture free energy.

Thanks for checking out my project and I look forward to seeing everyone's variations, suggestions and improvements!

Add Tip   Ask Question

About This Instructable
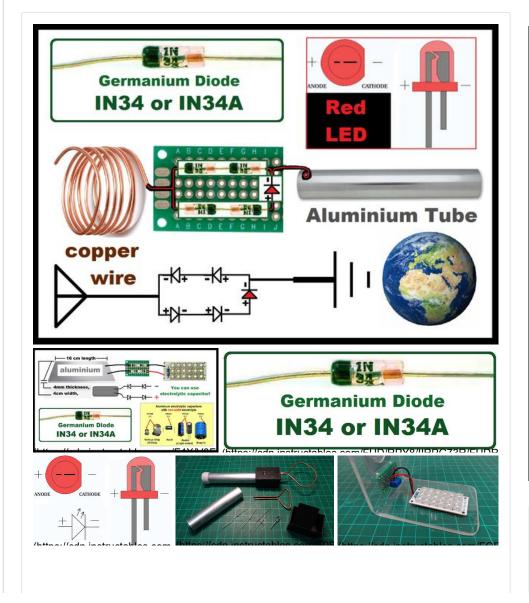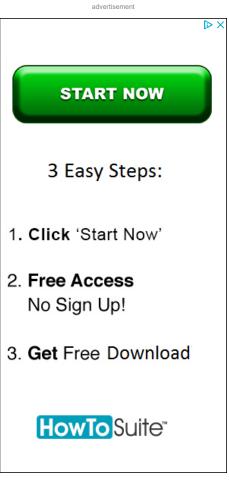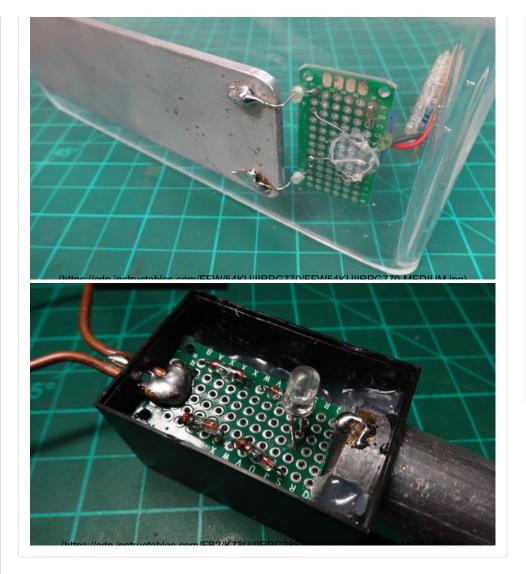
👁 **2,351** views

❤ **46** favorites

License:
(cc) BY-NC-SA

(https://cdn.instructables.com/FFW/54KU/IJRBG770/FFW54KUIJRBG770.MEDIUM.jpg)

(https://cdn.instructables.com/FR2/K73U/IJRBG78S/...MEDIUM.jpg)

How to Make a Electrostatic RF Detector or Ghost Detector, Hunting Solution. Much cheaper and easier than you may think For Under $5 ! A simple description is that IN34 or IN34A RF Germanium Diodes for Crystal Radio can take "Ambient" or "Radio Waves" to turns them into DC energy. I would like to show the Tesla Coil Transmitter, where I made an homemade "RF Detector" this is the simplest detector! without any batteries or hardships circuit, only with IN34 diodes and Ether to Earth!.. Stay Tuned and Enjoy! Please click the "like" button, subscribe my channel, and share this with your friends... To support my "work" every little bit helps! Watch the "ADS" THANKS

- See the original video https://www.youtube.com/watch?v=cSxqeGVC9t0
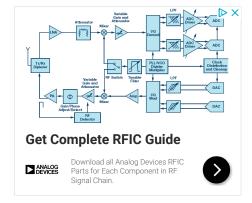
Step 1:

GEORGE CHANIOTAKIS (/member/GEORGE+CHANIOTAKIS/)
(http://https://www.youtube.com/
Follow 84
(/member/GEORGE+CHANIOTAKIS/)

Bio: A big hello to all mankind! This is my channel I hope you can find here various interesting things on in Renewable Energy Technologies and ... More » (/member/GEORGE%20CHANIOTAKIS/)

More by GEORGE CHANIOTAKIS:

(/id/MAKE-PERMANENT-MAGNET-SYNCHRONOUS-MOTOR-GENERATOR/)

(/id/Make-Wind-Turbine/)

(/id/High-Voltage-Die/)

Comments

We have a be nice comment policy.
Please be positive and constructive.

🖐 I Made it!   📷 Add Images   Post Comment
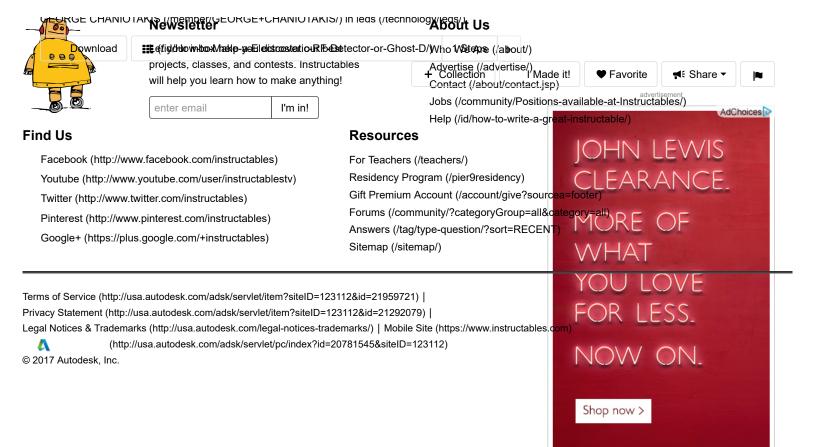
LazaroG4 (/member/LazaroG4/)   2017-11-14   Reply

I can't download and I was doing it before. I now need to go premium. This why i do not give any of my own files for the idea i have posted for free on You Tube. sorry for the not using of file and I have integrity and will not copy and paste as many do for free.

⬇ More Comments

How to Make a Electrostatic RF Detector or Ghost Detector   by

GEORGE CHANIOTAKIS (/member/GEORGE+CHANIOTAKIS/)

GEORGE CHANIOTAKIS (/member/GEORGE+CHANIOTAKIS/) in leds (/technology/leds/)

Download

(/id/How-to-Make-a-Electrostatic-RF-Detector-or-Ghost-D/)

Newsletter

enter email (/id/How-to-Make-a-Electrostatic-RF-Detector-or-Ghost-D/) ... projects, classes, and contests. Instructables will help you learn how to make anything!

enter email | I'm in!

Find Us

Facebook (http://www.facebook.com/instructables)

Youtube (http://www.youtube.com/user/instructablestv)

Twitter (http://www.twitter.com/instructables)

Pinterest (http://www.pinterest.com/instructables)

Google+ (https://plus.google.com/+instructables)

About Us

Who We Are (/about/)

+ Collection   I Made it!   ♥ Favorite   📢 Share ▾   🏳

Advertise (/advertise/)

Contact (/about/contact.jsp)

Jobs (/community/Positions-available-at-Instructables/)

Help (/id/how-to-write-a-great-instructable/)

Resources

For Teachers (/teachers/)

Residency Program (/pier9residency)

Gift Premium Account (/account/give?sourcea=footer)

Forums (/community/?categoryGroup=all&category=all)

Answers (/tag/type-question/?sort=RECENT)

Sitemap (/sitemap/)

Terms of Service (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21959721) |

Privacy Statement (http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=21292079) |

Legal Notices & Trademarks (http://usa.autodesk.com/legal-notices-trademarks/) | Mobile Site (https://www.instructables.com)

(http://usa.autodesk.com/adsk/servlet/pc/index?id=20781545&siteID=123112)

© 2017 Autodesk, Inc.

Build a Homebrew Radio Telescope

Explore the basics of radio astronomy with this easy to construct telescope.

Mark Spencer, WA8SME

There are many ham radio related activities that provide a rich opportunity to explore and learn more about the science of radio. One of those opportunities is radio astronomy.
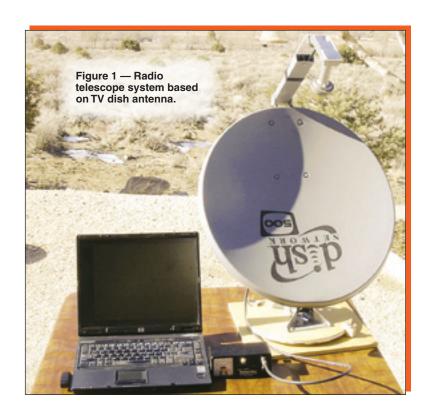
All matter emits radio frequency (RF) energy dependent on the temperature and makeup of the matter, including the matter in space. The foundation of radio astronomy is to study the heavens by collecting and analyzing the RF energy that is emitted by bodies in space, very much as optical astronomers use light energy collected by telescopes. It sounds complicated. While professionals use very sophisticated and expensive equipment, you can, with some simple equipment and a little investment, build a radio telescope that will allow you to learn and explore the fundamentals of radio astronomy.

A Homemade Radio Telescope

In this article, I will build on an existing design of a radio telescope made from one of those ubiquitous TV dish antennas that you see around your neighborhood. The radio telescope (RT) project described here can easily be reproduced. Although this is not a fully capable RT, it can provide a wonderful learning opportunity for you, or perhaps students in your local school.

Figure 1 shows the radio telescope set up. The major components include a modified TV dish antenna mounted on a wooden support structure to allow pointing the antenna, a commercial satellite signal strength detector that displays the signal strength of signals collected by the dish on a meter and an interface that converts the signal strength into a amplitude modulated tone. The tone is fed into a computer sound card and finally a computer and software graphically displays the signal strength as a function of time.

The TV dish modifications are structural, and any available TV dish system can be used. The signal strength detector costs between $40 and $65 and is widely available from Web retailers. The interface circuit, which will be described shortly, is easily duplicated and costs approximately $20. Finally, the display software is free.

Figure 1 — Radio telescope system based on TV dish antenna.

Figure 2 — Dual LNB mount. Note two coax connectors.

Figure 3 — Homemade plastic single LNB mounting bracket.
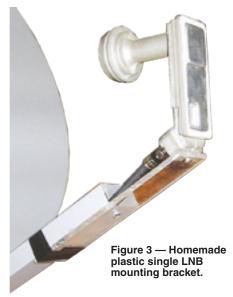
What it Can Do

The following is just a sample of what you can do with this simple RT:

- Use the sun to study and determine the beamwidth of the dish and verify the mathematic formula that is used to predict dish antenna performance.
- Measure the radiation intensity of the Sun and perhaps detect changes in solar activity.
- Measure the relative changes in the surface temperature of the moon.
- Learn about and explore a common radio astronomy collection technique called the *drift scan*.

Figure 5 — CM satellite signal strength meter.

Figure 6 — *SkyPipe* screen showing antenna response.

Figure 4 — Dual coax connector configured LNB. Terminate one connector with a dummy load.

- Explore the fundamental principle of energy emission as a function of temperature by detecting the relative differences between the temperatures of emitting bodies.
- Detect satellites parked along the Clarke Belt in geosynchronous orbit and illustrate how crowded space has become.
- Detect the Earth's rotation around the Sun and the Earth's spin on its axis by comparing daily drift scans of the horizon.

Antenna Subsystem

The basic RT system is based on the "Itty-Bitty" design that is described in two Web pages.[1,2] The TV dish is an offset 18 inch dish that has down converter(s) mounted at the focal point of the dish. The down converter is called a low noise block (LNB). The LNB is a preamplifier/down converter that converts the satellite signals from around 12 GHz down to around 2.4 GHz. Most modern dishes have two or more LNBs to access more than one TV satellite at a time without changing the pointing of the dish

[1]Notes appear on page 45.

(Figure 2). The LNBs are mounted to share the focal point of the dish. Since only one LNB is required for the RT, I made a minor adjustment to the published Itty-Bitty design to position the single LNB at the dish focal point. Mounting the single LNB at the focal point really helps in pointing the antenna.

I used the existing LNB housing and mounting bracket as a template to determine the distance between the edge of the mounting arm to the mounting hole of the LNB. I then used a piece of plastic to fabricate a new mounting bracket for the LNB as shown

Figure 7 — RT Interface circuit diagram.

Figure 8 — RT Interface block diagram.
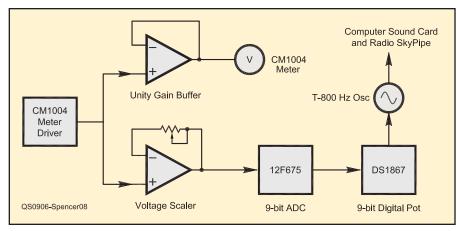
Figure 9 — Power and ground connection to CM board.

in Figure 3. The dimensions are not super critical, but careful placement certainly will improve the RT performance.

Some LNBs have two coax connectors. Only one will be used in the RT (Figure 4). It is a good idea to terminate the extra coax connector with a 75 Ω dummy load plug to balance the load on the LNB. The dummy loads for F type TV coax connectors are readily available from electronic parts retailers.

Note that the dish is mounted upside down. Though this orientation is not ideal for receiving satellite signals, this arrangement helps with pointing the dish in its radio telescope role.

Satellite Detector

The detector used in this project is the Channel Master (CM) satellite signal level meter model 1004IFD (Figure 5).[3] The CM is connected to the LNB. Power is supplied to the LNB through the coax connection from the CM. The CM detects the signal coming from the LNB and gives a meter indication of the signal strength and also varies the frequency of an audio tone to help technicians point the dish at the desired satellite. As you move the dish through the beam coming from the satellite, the meter indication will increase and then decrease coincident with the pitch of the audio tone.

The Itty-Bitty plans detail how to connect power to the CM and in turn connect power to the LNB (this power connection is handled by the interface in this project). Though somewhat effective, the CM meter and variable frequency tone indications provide limited utility in detecting changes in signal strengths required for radio astronomy.

Display

To really study the signals received by the RT, you will need to see them displayed graphically on a strip chart. There is an excellent software package called *Radio-*

SkyPipe that is posted on radio astronomy Web sites.[4] The free version of this software is a good place to start. *SkyPipe* uses the computer sound card to measure the incoming signal strength and graphically displays the signal strength as a function of time. Figure 6 is illustrative of a signals detected by the RT. *SkyPipe* is very easy to use but some study of the HELP files will make it easier for you to fully tap into the capabilities of this software.

SkyPipe requires audio signals to be fed into the sound card MICROPHONE jack. The output of the CM detector is either an analog meter reading or a frequency modulated (constant amplitude) tone that is not really compatible with *SkyPipe*. An interface is required.

Interface

What is required to make the CM output work with *SkyPipe* and a sound card is to convert the signal level into an amplitude varying audio tone. The interface designed to do this is shown in Figure 7 and as a block diagram in Figure 8. Refer to the block diagram during the description of the interface function.

The unity-gain op-amp is used as a buffer between the CM meter driver circuit and the analog meter. The other op-amp is used as a voltage multiplier to scale the CM meter driver output voltage to match the 5 V reference voltage of the following analog
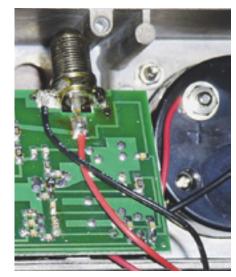
to digital converter (ADC). The variable resistor in this voltage multiplier circuit is used to calibrate the CM to *SkyPipe*. The voltage from the multiplier is fed to a programmable interface controller (PIC) that is programmed as a 9-bit ADC to covert the analog voltage that is a function of received signal strength to a 9-bit digital word that is used to control a digitally controlled variable resistor. The interface includes a simple Twin-T audio oscillator circuit that provides a tone of approximately 800 Hz that is fed to the computer sound card. The amplitude of this audio oscillator is varied by the digital pot that is being controlled by the PIC. The result is the audio amplitude being varied in step with the signal strength detected by the CM.

The circuit provides power to the CM and the LNB. A 12 V source in the CM is tapped through an RF choke and this is connected to the LNB coax connector inside the CM (Figure 9). The 12 V is also regulated to 5 V to provide power to the interface. Though probably not required, there are two 5 V sources, one for the digital components of the interface, and the other for the analog components with one common ground point. This arrangement is used to isolate potential digital and analog noise sources within the circuit.

The interface is built on a circuit board and mounted right inside the CM box

Figure 10 — CM with interface board.

Figure 11 — Aiming the RT at the Sun, note LNB shadow location.

Figure 13 — Drift scan of the Sun indicating antenna's azimuth pattern.

(Figure 10). Though I made an etched circuit board for the circuit, the hand wired prototype worked equally well for those who would rather roll their own. The PIC firmware is available on the *QST* Web site.[6]

RT in Action

The first thing you need to do is learn how to point the RT antenna. The best place to start is to connect the CM to the antenna and point the antenna at the Sun. *Caution*: Do not look into the Sun as you do this, or at any time. Adjust the pointing angle and elevation until you get peak signal strength as indicated on the CM meter or hear the highest pitch audio tone. With the antenna pointed directly at the Sun, take note of the position of the shadow of the LNB on the surface of the dish (left in Figure 11). If you look from behind the dish, along the LNB
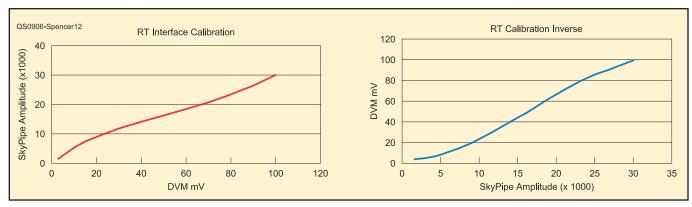
Figure 12 — Example calibration curves.
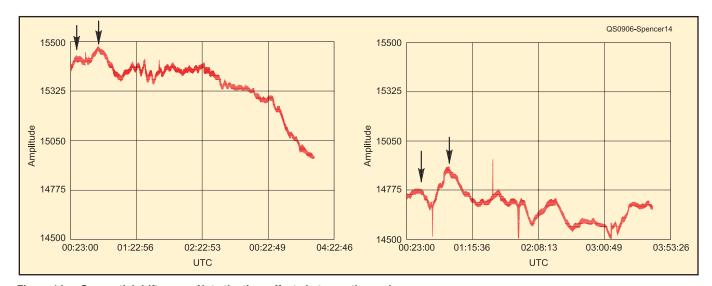
Figure 14 — Sequential drift scans. Note the time offsets between the peaks.

From June 2009 QST © ARRL

supporting arm (between the arm and the rim of the dish), you will see the Sun being blocked by the LNB.

Once you have the RT set up, it needs to be calibrated to match the output of the CM to *SkyPipe*. I have developed an *Excel* spreadsheet template to help with the calibration and a few of the other activities that you can accomplish with the RT (also available from the *QST* Web site). Turn the RT to a signal source, the Sun, or the side of a building would work. Turn the gain control of the CM to set the meter to maximum. Run *SkyPipe* and adjust the variable resistor on the interface board until you get a reading on the *SkyPipe* graph vertical (y) axis of approximately 32,000. With the maximum value set, adjust the CM gain control through the voltage range (0 to 100 mV) in 10 mV steps and record the corresponding y axis value on *SkyPipe*. This data is entered into the *Excel* spreadsheet to compute the calibration curve between voltage and y axis value. Both voltage and y axis values are used in analyzing recorded signal strength data (Figure 12).

A good first activity is to do a drift scan of the Sun. A drift scan means that you set the antenna azimuth (AZ) and elevation (EL) to some fixed pointing angle and allow the Earth to serve as the rotator to drag the antenna across the sky. To do a drift scan of the Sun, first set the elevation and azimuth to point directly at the Sun (maximum signal) and then move the azimuth toward the west
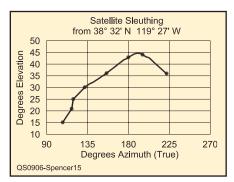
Figure 15 — Clarke Belt plot — tracking down satellites.

(leave the elevation set) until you are off the peak signal. Now start *SkyPipe*. In about 15 minutes, the Sun will pass through the antenna pattern beam width and the result will be as illustrated in Figure 13. You can also use this collection technique to explore the antenna performance parameters.

A good second activity is to do two drift scans of the night sky on two consecutive nights (beginning the scans at the same time each night) using the same fixed antenna azimuth (AZ) and elevation (EL). Figure 14 shows two such drift scans. Although at first glance they may not seem similar, there are some interesting features that are pointed to by arrows. If you compare the time that these two peaks occurred, the time difference is about 4.5 minutes. This shift is the result of the distance the Earth had traveled during the 24 hours between collections.

This illustrates that the Earth's rotation as well as its travel in orbit needs to be considered when comparing drift scans. Enough to make your head spin (pun intended)?

A final good starting activity is to point the antenna toward the Clarke Belt and find all the satellites in geosynchronous orbit transmitting on 12 GHz. If you record signal strength peaks and AZ and EL for each peak, you will develop a graph of the Clarke belt as illustrated in Figure 15.

I have only scratched the surface, and the sky is the limit of this little project. The RT project can certainly broaden your horizons and expand your understanding of our universe. If you would like more detail than can be presented here, please contact the author.

Notes
[1]www.setileague.org/articles/lbt.pdf.
[2]www.aoc.nrao.edu/epo/teachers/ittybitty/
   procedure.html.
[3]www.pctinternational.com/channelmas-
   ter/0612/satellite.html.
[4]radiosky.com/skypipeishere.html.
[5]en.wikipedia.org/wiki/Geostationary.
[6]www.arrl.org/files/qst-binaries/.

Mark Spencer, WA8SME, is ARRL Education and Technology Program Coordinator. He is an ARRL member and holds an Amateur Extra class license. You can reach Mark at 774 Eastside Rd, Coleville, CA 96107 or at **wa8sme@arrl.org**.

Simple demonstration to explore the radio waves generated by a mobile phone.

Dr Jonathan Hare, Sussex University, Department of Physics, Falmer, Brighton. BN1 9QH

Also see the brief full article at: <u>mobile phone detector</u>

IMPORTANT NOTE: this device works very well on the old style mobile phones (as shown in the photo above). However, it does not always work well with modern smart phones. This may be because modern phones use higher frequencies, less power and use the power in a slightly different way (e.g. spread spectrum). Some smart phones do work and success may be due to the signal strength of the local mobile phone mast nearby. If you are in a low signal area the phone will create more power to ensure reliable communications. If you are in a very strong signal area (very near the local network) your phone will drop its output power and consiquently there will be less power to pick-up and to convert to a voltage to light the LED.

Described is a simple low cost home-made device that converts the radio wave energy from a mobile phone signal into electricity to light an LED. No battery or complex circuitry is required. The device can form the basis of a range of interesting experiments on the physics and technology of our mobile phones.
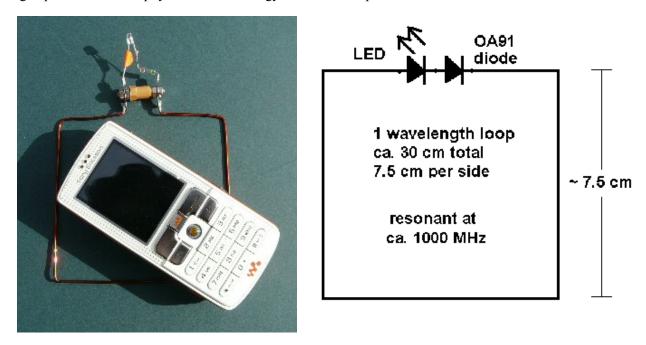
Fig. 1: left: mobile phone radio wave detector and right: the simple schematic

Introduction

Electromagnetic radiation (EMR) is at the heart of modern mobile phone data communications networks. The way a mobile phone and local base stations (the antenna covered masts you see dotted all around the place) communicate between each other is by using EMR in the radio wave part of the spectrum [1,2,3]. On switch-on your mobile sends digital information pulses by rapidly switching on and off the radio waves rather like a fast Morse code signal. Your text or voice is also converted into a series of digital pulses and sent across the network to be decoded (reassembled) by another mobile phone you dialled.

EM radiation and radio waves

Mobiles make use of various bands of radio frequencies to communicate between the mobile to base and the base to mobile: in Europe these include 900 and 1800 MHz (850 and 1900 MHz in the USA and Canada) [2, 3].

The relationship between wavelength, speed of light and the frequency follows the well known formula:

Wavelength λ (m) = speed / frequency = c (ms^{-1}) / ν (Hz)

λ (m) = 300,000,000 / ν (Hz) or approximately:

λ (m) = 300 / ν (MHz) ..... Equation 1.

So for a mid-range of about 1000 MHz (1 GHz) we get a typical mobile phone wavelength of about:

λ = 300/ 1000 = 0.3 m = 30 cm.

Simple radio wave detector

The loop consists of about a wavelength of wire, ca. 30 cm so each side is about 30/4 = 7.5 cm. The dimensions are not critical. The two ends are connected directly to a simple series circuit consisting of a high brightness LED and a germanium diode. They need to be connected correctly. All these components are cheap and readily available from electronic stores [4]. The loop can be made from a piece of copper wire roughly bent into a square (although a circular loop or rectangle will also work). If the wire is insulated remember to scrap off the insulation and solder-tin the ends. Simply solder the germanium diode and LED into circuit as shown in the diagram.

On a new LED the long lead is the positive (anode) while the short lead is the negative (cathode). The germanium diode has a line (band) around the end which is the cathode. When correctly wired the LED and the germanium diodes are connected so they both allow current to pass in the same direction, i.e. in the circuit diagram the arrows point in the same direction. In practice this means the LED and germanium diode are joined at the cathode of one and the anode of the other.

In my prototypes I used an insulator between the loop ends (light coloured cylinder in the photo) to make the whole thing more sturdy but this was purely for mechanical reasons and is not needed for the circuit to function properly. Note: a much more sensitive version using a x10 and x100 DC amplifier is described on my web site [6].

How it works and how to use it

When a radio wave passes across a metal object the EM fields cause the charged electrons in the metal to oscillate and this causes small AC currents at the same frequency to be induced into the metal. If a mobile is brought near to the loop and a call or text is made [5] the radio waves emitted from the phone pass across the loop. This induces a voltage into the antenna (the loop) and if it is close enough will be large enough to light the LED. As the loop is about one wavelength in size it is resonant and so there is a good transfer of power (low reactance) between the radio wave and LED.

The mobile phone automatically tests the network and adjusts its transmission power to maximise the battery life and minimise network interference. As a result the brightness of the LED will depend on the data being sent (the average signal), the local signal strength and how close the loop is to the phone. Why the second diode? - It's curious why the germanium diode is needed at all. The LED is a Light Emitting Diode after all and one would not think that another diode would help. However my initial experiments failed because I had not included it. The LED will have a relatively high capacitance which at these frequencies will tend to de-tune the loop and short out the LED. The germanium diode however is made up of a tiny wire which only makes a point-of-contact onto a piece of semiconducting germanium so it's 'self' capacitance is very low keeping the loop resonant.

The germanium diode will rectify the AC signal from the loop forming a series of DC pulses that will be nicely smoothed by the LED's capacitance. Without the diode however the raw AC signal from the loop will tend to be averaged to zero by the LED's capacitance.
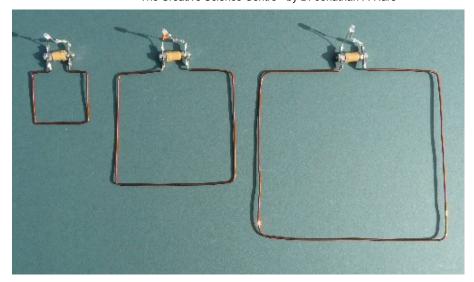
Fig. 2: Three loops in a row of varying size. The one described here and shown in Fig. 1 is shown in the middle. Smaller ones may well work better for higher frequencies such as the 3G networks (see below).

Other size loops
Fig. 2 shows a set of three loop devices with edge lengths of roughly 3.7, 7.5 and 15 cm. You can find out for yourself that the best match to the mobile signal is with the full wave loop of ca. 7.5cm per side. The other loops do work to varying degrees however (smaller ones may work better for the 3G network). The larger loop works well for the '70 cm' amateur radio bands.

Polarisation
The electric and magnetic fields making up the EM wave are orthogonal (they are at right angles to each other as they pass through space) to each other but depending how they are generated by the transmitting antenna can arrange themselves in any orientation with respect to the ground. If the electric field is parallel with the ground we say the wave is 'horizontally polarised' while if its normal to the ground we say its 'vertically polarised'. The loop antenna will respond best to one type of polarisation (depending on its orientation) so it's worth experimenting with the orientation of the mobile (or the loop) to get the strongest signal - brightest LED.

Mobile antenna
Inside your mobile phone is a transmitter / receiver and antenna. Many mobiles have this antenna at the top of the phone but some of the PDA type phones have it at the bottom. As a result you can locate the position of the antenna by moving it around the center of the loop till you get maximum LED brightness.

Networks
There are various different networks that a mobile may use both in the UK and abroad. It may be that you need to adjust the network phone settings on your mobile i.e. change from "automatic select" to set for "GSM" so as it get the strongest signal to light the LED. Note: the 3G network might not be powerful enough to light the LED. As the GMS network is currently the main network over the UK the device should work anywhere where you can get a signal as long as you check the correct selection on your mobile menus [5, 7]. The 3G network operates on a higher frequency (smaller wavelength) so you might find a smaller loop will work better than the main one described here. See 'other experiments' section below.

Test signals
In order to pick up the radio wave energy from the phone it obviously needs to be transmitting a signal. There are a few ways to do this:

1) On switch 'on' (or change of network) you can see that the mobile initially transmits for a few seconds to the network to tell it it's there (especially if you have moved since turning it off). You don't actually need to dial a number to detect these signals.
2) Even if don't text or call, throughout the day the mobile will send out data to 'keep up' with the network, especially if you are moving around (going through train tunnels etc. see below).

3) When you make a phone call you will transmit. Initially there is quite a lot of data being sent but in a few seconds data / power only gets transmitted when you speak. So to light the LED continuously you need to talk or provide some background sound continuously. Your service provider voicemail might be a good free phone number to try for these experiments [5].

4) Texting is the easiest way to show the radio wave power being transmitted. Long texts will light the LED for longer than short texts.

5) Finally set up the mobile on the loop and use another phone to text or phone the mobile. Even though you are not directly using the phone you will see that even on 'receive' the mobile phone transmits data to and fro. Ring off before you get charged.

Note: If you can use a free phone number it will save you money [5].
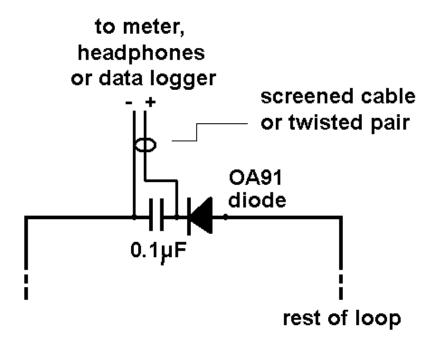
Fig. 3: Adding a capacitor and coax (or twin) lead so that headphones, a meter or a data logger can be connected (Note: diode is reverse wired compared to Fig. 1).

Other experiments:
Hearing data - if headphones are wired across the LED they will convert the voltages into sound and you can 'hear' the clicks of the digital data being transmitted. These are the same clicks that so easily get picked up by sensitive electronics such as a stereo amp or recording equipment when making a video for example. Hence - 'no phones on' when filming.

Logging data - if a meter, or better still a stand-alone data logger, is attached across the LED then one can monitor the EMR from the phone. For example even if you are not making a call your mobile will send signals too (and receive signals from) the network while travelling around. Fig. 3 shows a simple modification using a de-coupling capacitor so that a coax cable (or twisted pair) can be used to go to headphones, meter or data logger. Note the diode has been reversed so that the logger has the correct + and – connections for a unipolar input logger. The capacitor should help average the signal and stop radio frequencies going down to the logger. If one is available a few turns of the wire can be wound within a ferrite ring near to the logger so that maximum immunity to the mobile phone signal can be obtained for the logger electronics.
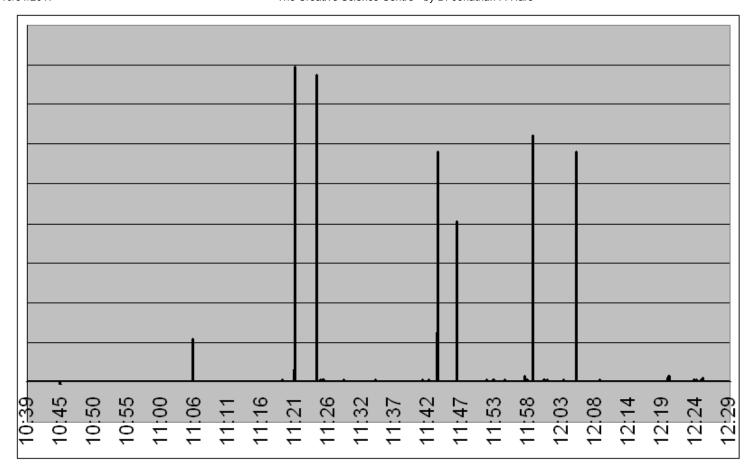
Fig. 4: Typical mobile phone data signals sent out onto the network while travelling around. These were recorded by a data logger from a mobile using the loop (no calls or text were made) while travelling on the train from Brighton to London Victoria (and then around London and return). Many of these peaks were the phone sending out 'I am here' data after coming out of one of the many long tunnels under the South Downs during the journey.

Out and about - Once you can log data you can discover all sorts of interesting things your mobile phone is doing without you realising it. Fig. 4 shows the plot over a few hours of travelling between Brighton and London (and within London) on the train. The detector was simply placed near to a phone that was not making or receiving a phone call or text, but was turned on.

The graph shows that the mobile sends out signals to tell the network where it is as it travels along and in particular goes in and out of long train tunnels. The peak heights vary because of the different powers the mobile transmits at depending on the signal strength of the local network and also because of the way the data logger 'snatches' a reading from the circuit every few seconds. As your phone sends out data onto the network to ensure the very best communications as you move around, so your mobile and the network obviously knows where you are and where you have been. Thieves and criminals beware the police can track you!

The inverse square law - If the transmitting mobile phone is moved away from the loop one would expect the signal to drop off. Unfortunately because both diodes need a certain threshold before they conduct the detector is not sensitive to small signals and not very linear. Therefore it's not very easy to use the device to measure the inverse square law (drop in signal v distance away) but of course you can see the signal go down. You could perhaps use the device to plot isobars - i.e. plot the equal intensity signals around the phone / nearby objects.

Changing the resonant frequency of the loop - you might be able to make some simple sliding mechanism (e.g. a small trombone-like mechanism) out of metal tube for example to tune the loop device for different frequencies. Then you can use it to find the average wavelength and so determine the center frequency by adjusting the size for maximum brightness of the LED. The wavelength can be determined by measuring the total distance around the loop. If we assume the antenna is one wavelength in total length then the frequency can be established by rearranging Equation 1, i.e. v (MHz) = 30,000 / L (cm), where L is the length around the loop (cm).

Note: You will need to allow the transmitted digital signal to 'settle down' i.e. make measurements only after a few seconds after dialling / pick up so that only the sound data is being transmitted rather than the initial connection data. A constant sound will also need to be made so that the mobile phone continuously transmits data. It's worth playing music near to the phone or constantly whistling to keep sound coming into the phones microphone.

Mobile phone detector - teachers who want to know if the students / pupils really have turned-off their mobile phones (rather than just put on 'silent') can wire the loop device into the class room white-board speakers. Any mobile that is on in the class will send out signals which (if you are close enough) you will hear the data going to and fro - you will have your very own 'who's got their mobiles on' device which might be useful for exams etc.
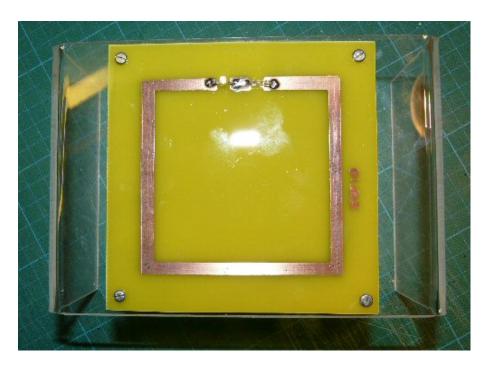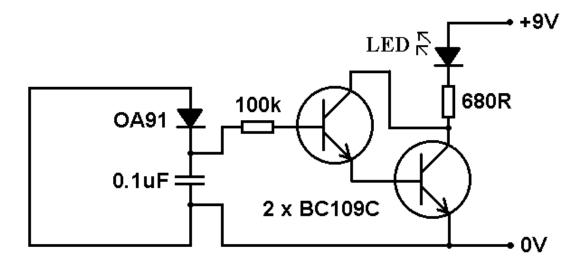
Fig. 5: The SEPNet 'deluxe' printed circuit board version (pcb) on a perspex stand where the loop is composed of a pcb copper track and the diode and LED soldered onto the board (top) [8,9].

Summary
All in all then, for such a simple easy to make device I hope you agree that there is a lot of scope for interesting science / technology investigations with your mobile phone. The device would make a good science week project (for radio amateur clubs etc.) A 'deluxe' pcb version (Fig. 5) on a perspex display case (Fig. 5) is currently going around the southern UK as part of the SEPnet outreach work, see the 'Radiation Exhibition' [8] and also as part of my on-going lecture series [9].

Post publication additions
(What follows was not included in the published article as this calculation was worked out later). A full wave loop is resonant and so looks purely resistive to the radio waves. Such a loop will have a resistance of about 100 ohms (Note: this is the AC resistance and not the DC resistance which will be very low). Now power $P = V \times I$ (V = voltage and I = current) and resistance = $R = V/I$ therefore $P = V^2 / R$ or rearranging $V = \sqrt{(P \times R)}$ which means that the voltage created by a power level of say 50mW (say for argument that roughly half the mobile phone power) arriving at the antenna will be about $V = \sqrt{(100 \times .05)}$ which is aprox. $V = 2V$, enough to light an LED.

This circuit uses a two transistor darlington driver to amplify the signal from the loop and diode making the detector much more sensitive. The LED will be much brighter using this circuit. Note: the circuit needs a battery to power it (e.g. a PP3 9V)

References
[1] These ultra high frequencies (UHF, > 1000 MHz) are also often called microwaves.
[2] wiki pages
[3] Elektor Electronics magazine, June 2005
[4] order codes for the germanium diode and LED are:
e.g. Germanium diode: Maplin Electronics: QH71N, Rapid Electronics: 47-3114
e.g. LED: Maplin Electronics: UF72P , Rapid Electronics: 55-0085
[5] to save money use your voice mail service (often you simply dial 121).
[6] for details of an amplified detector see: wavemeter
[7] select 'network setting' from the mobile phone 'settings' menu and then go to 'network mode' and select 'GSM 900/1800' rather than 'automatic'.
[8] SEPnet mobile phone device on display throughout southern UK. [9] for details of my talks see: talks and workshops

THE CREATIVE SCIENCE CENTRE

Dr Jonathan Hare, Brighton, East Sussex. BN1 9QJ.
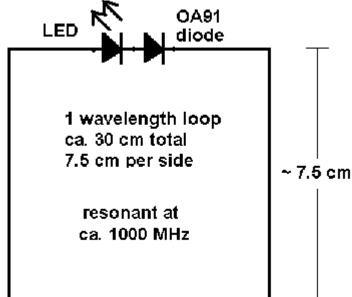
home | diary | whats on | CSC summary | latest news

Simple demonstration to show mobile phones emit radio waves

Dr Jonathan Hare, Sussex University, Department of Physics, Falmer, Brighton. BN1 9QH
Note: this article is in press: Elektor Magazine, July-August 2010, p. 56-57

For other experiments with this device please see my full article at: mobile phone detector

left: mobile phone radio wave detector and right: the simple schematic. Below: detail of the LED and germanium diode.

IMPORTANT NOTE: this device works very well on the old style mobile phones (as shown in the photo above). However, it does not always work well with modern smart phones. This may be because modern phones use higher frequencies, less power and use the power in a slightly different way (e.g. spread

spectrum). Some smart phones do work and success may be due to the signal strength of the local mobile phone mast nearby. If you are in a low signal area the phone will create more power to ensure reliable communications. If you are in a very strong signal area (very near the local network) your phone will drop its output power and consiquently there will be less power to pick-up and to convert to a voltage to light the LED.

This is a very simple and cheap device that demonstrates mobile phones ('cell phones' or 'handies') generate radio waves. We have a 30 cm (7.5 cm per side) full-wavelength loop antenna (a 'Quad' to radio amateurs) connected to a germanium diode and a hyper-bright LED. The loop can be made of copper wire, thin sheet metal or a track on a pcb. The diodes need to be wired correctly. I think the germanium diode is needed as the LED probably has too great a self-capacitance to perform at the very high AC frequencies generated by the phone (ca. 900 or 1800 Hz) but will work well with the DC pulses from the germanium diode (which has a very small capacitance).

To show the mobile generates radio waves put the mobile near to the loop and dial a number (use a free phone number, e.g. your voice mail) or text. The radio waves will induce a voltage into the loop, large enough to light the LED. The LED will flash indicating the digital data being sent by the mobile phone transmitter. You may need to set your phone to 'GSM 900/1800' rather than the '3G' network in the settings menu.

parts:
germanium diode: Maplin Electronics: QH71N or Rapid Electronics: 47-3114
LED: Maplin Electronics: UF72P or Rapid Electronics: 55-0085

A very simple connector block version and a circular 1 wavelength loop

THE CREATIVE SCIENCE CENTRE

Dr Jonathan Hare, Brighton, East Sussex. BN1 9QJ.

home | diary | whats on | CSC summary | latest news

BACK TO G1EXG RADIO PAGE

EPAD MOSFET - 'near zero' and 'zero' threshold devices
(e.g. the ALD110900)

I will add more details to this page later on but for the moment it gives links to important and inspiring sites.

ARTICLES
1) High Sensitivity Crystal sets, Technical Topics, Pat Hawker, Rad Com, p. 77-78, May 2007.
2) A novel kind of 'crystal set' radio, Giles Read, Rad Com, p. 60-61, June 2007.
3) also see Pat Hawker, Technical Topics, Rad. Com. January 200, page 53-57.
4) High Sensitivity Crystal Set by By Bob Culter (N7FKI)
5) Next-generation Zero-Threshold Voltage EPAD™ design enables circuits with greater operating range in low voltage supply environments
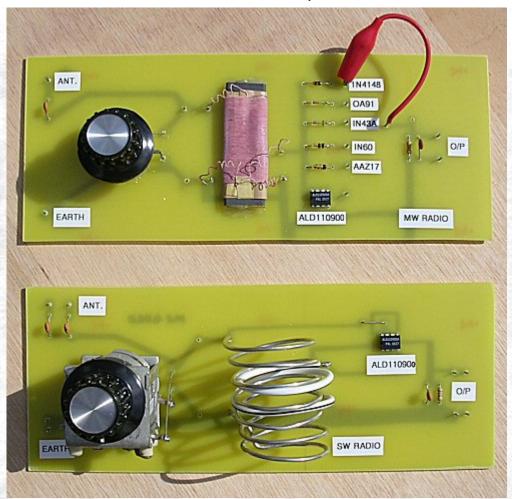
LINKS ABOUT THE CHIPS / DEVICES
1) Advanced Linear Devices web site
2) EPAD MOSFET (ALD web site)
3) ALD110800 and ALD110900 page (ALD web site)
4) ALD110802 and ALD110802 page (ALD web site)

ORDERING THE CHIPS / DEVICES
1) mouser.com (USA)

'CRYSTAL SET' RADIO'S - DIFFERENT DETECTORS

**Top: the medium wave (MW) test radio with a selection of 'detectors' (selected by the croc clip lead)
Bottom: the short wave (SW) version with a ALD110900 detector
(on each circuit board the connections are: antenna - top left, Earth - bottom left, audio output
(heaphones / amp) - far right.)**

RADIO AND RADIO WAVES

The simplest 'radio' can be a piece of wire attached to the input of an amplifier. Why does this pick up radio signals? Well in principle it should not be able to pick up anything but in reality poor solder joints in the amplifier circuitry, as well as point contact effects between the ends of the anenna wire going into the input socket of the amp as well as other effects mean that you do often hear radio signals.

Radio waves are electromagnetic waves as they pass through a metal they induce small voltages into it. Any piece of metal e.g. metal fram specs, tape measures, metal window frames, a piece of wire etc. will act as an antenna and have tiny voltages induced into them. These voltages will be due to natural radio waves (from Space as well as the Sun and Earth), radio stations, satellites, mobile phones, garage door remote controls, microwave ovens .... the list goes on and on.

In the case of long, medium and short wave radio stations the amplitude of the radio wave signal is modified by the music, voice or program - we call it Amplitude modulation AM. Here the strength of the radio waves varies as the tones, loudness and pitch of the program vary. As a result the voltages induced in the metal object antennas distant from the radio wave stations also vary in accord this program information or modulation.

To actually hear the programs on the radio waves you cant actually take this tiny signal and listen to it directly (say with headphones or an amplifier). This is because the signal is a very high frequency signal out of the range of human hearing. To get the audible information - the program - 'off' the radio signal you need a device known

as a 'detector'. This is usually a diode but lots of other things can act as ineffecent detectors for example a mineral called galena, coke (burnt coal) ... as well as rusty screws and bad solder joints which is how the ampilfier mentioned above apparently picks up radio signals.

In the early days of radio crystals of galina were used as diodes and so these radios became known as 'crystal set radios'. They did not have any amplifiers and did not require any battery to work. They got all their power from that induced into the (of very long) antenna by the radio signal(s). The radio detector is an extremely important part of the radio and its proper function determines how well the radio works.
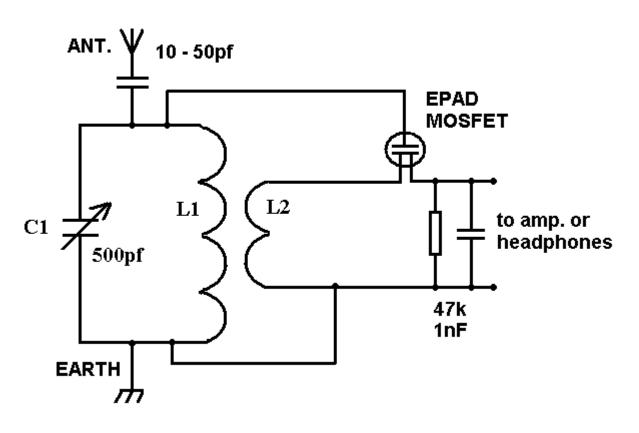
The radio signal is an AC signal. A speaker or headphones can not resonate at such high frequencies as radio waves and if you try to wire in the RF to the headphones the diaphram just averages the power. As half the time the RF is positive and half the time negative, in an AC signal, the average is zero and so you dont hear anything. The diode or detector is a device called a diode which only conducts electricity in one direction. So it only allows the positive, or negative half (one half but but not both) of the AC signal to go through (depending on which way it is wired). After the diode the average is now no longer zero its actually a changing signal dependant on the modulation of the radio wave - which is the information we want to hear. As this is at audio frequencies, and as the diaphram can move at audio frequencies, you hear the information on the detected RF signal - the music, voice or what ever.

Why should some detectors (diodes) be 'better' than others? The answer is not that some are better at 'magnifying' the signals but rather that each diode requires a certain threshold voltage in order to start conducting (one way). Signals below this level will not get passed through and a signal only slightly larger will therefore only pass through weekly (the diode unfortunatly absorbing the majority of the signal). As a result for large signals you wont expect to see much difference between different detectors but for weak signals there can be a great deal of difference. In an ideal world the diode would start to (forward) conduct at just above zero volts and this ideal detector (which does not conduct at all in reverse) would produce the greatest signal possible for the particular strength signal applied to it.

Recently I was playing around with OA91 germanium diodes for a mobile phone detector. I made a simple 10cm per side square loop and wired it directly to an LED. I was hoping that if I brought the mobile near to the loop and texted, or made a phone call, the RF should be picked up by the loop (which is almost resonant) and the LED should light - it didnt work! But putting an OA91 in series I did get it to work. So why did I need the OA91 diode when the LED is itself a diode?

It can't be due to the threshold voltage in this case as I am still using the LED and I did get it to work with the same signal! So I guess it because the capacitance of the LED is very high compared to the point-contact germanium diode. If we extend this thinking to the crystal set radio I guess the germanium diodes have very little capacitance and so dont allow any of the 'wrong' half of the AC signal through and so act as a more perfect detector. An LED would have a high capacitance and so let some of the wrong half of the AC signal through which on averaging by headphones would be nearer zero (see above) - hence less sound.

A low threshold (turn-on) voltage and a low self capacitance are crucial things in a good detector. Point contact germanium diodes having a very small area (a point of contact) have low capacitance and the germanium semiconductor-metal junction of the contact has a low turn on voltage. EPAD Mosfets also have low capacitance, high input resistance (low loading of the signal) and low turn on voltage.

MEDIUM WAVE TEST CIRCUIT (top in photo)

As a simple test circuit to compare detectors I made up a simple 'crystal set' radio from a ferrite rod coil from an old radio, tuning capacitor an array of diodes (crystals) and a simple RF filter composed of a 47k resistor and 1nF capacitor. The coil and capacitor formed the resonant circuit which was fed via a 10 - 50pf capacitor by the antenna (long wire). This capacitor helps to reduce the effect of the antennas own capacitance and inductance from modifying the resonant frequency of the tuned circuit. Ideally this should be as small a capacitance as you can get away with but too small a value will reduce the radios sensitivity ....

An earth was connected to the ground connection of the circuit. Instead of taking the antenna end of the resonant circuit to a diode detector (as many simple crystal sets do) I used a ca. 10 turn coupling coil around the coil. This was done to limit the loading on the resonant circuit by the detector. It should help keep the tuning sharp and also help to keep the overall (band) spread of tuning as large as possible.

Because of the reduction in turns (ca. 100 turns : 10 turns, i.e. 10:1) the coupling coil reduces the signal that is available to drive the detector. For example if I use a standard silicon diode (e.g. a 1N4148) which requires about 0.6V to conduct then we might need ca. $10 \times 0.6 = 6V$ of RF to be developed in the resonant circuit! A high Q resonant circuit might do this with a long antenna coupled to it but without such a generious antenna such a set-up wont be very sensitive.

RESULTS

I tried an array of 'crystal set' type diodes to campare them. Typical rough results are shown in the table below. This was with an Earth and about 3-4m of antenna wire randomly strung near to ground level (so its a pretty bad set up really - a good test).

The EPAD MOSFET data sheets described how the gate and drain of one of the MOSFETS can be joined together to form the anode of a diode the source then forms the cathode. I tried this arangement with the ALD110900 in order to make up a 'crystal set' diode from this MOSFET but it didnt perform well.

Then I tried the alternative arrangement as described in the references above. In these articles they dont use the MOSFETS as diodes but connect the gate directly to the resonant circuit. As the input resistance is extremely high (c.a. 1E12 ohms) and the input capacitance is very low (c.a 3pf) it does not load the tuned circuit by any appreciable amount. The source and drain are then wired between the coupling coil and filter just as the usual diode detector is in a crystal set radio. In this configuration the MOSFET is working more as a gate voltage controlled resistor rather than a diode i.e 'off' for half an RF cycle and 'on' for the rest - this lets the envelope through and you hear the music or sound. One article describes the arrangement as a synronious detector. As there are two MOSFETS in one package the two are wired in parallel i.e. gate to gate, drain to drain (there is a common source) etc.

DETECTOR TYPE	AF voltage produced after filtering
1N4148 silicon	no signal
OA91 germanium	1mV
1N43A germanium	1mV
1N60 germanium	1mV
AAZ17 germanium	1.5mV
ALD110900 EPAD MOSFET wired as diode	<1mV
ALD110900 EPAD MOSFET wired as sync. detc.	10mV

Typical results of the AF voltages produced by the detector (for use by headphones or amplifier) for an Earth connection, small (ca. 3-4m antenna wire near ground level) and the prototype 'crystal set' radio.

For the simple set-up described above there was not enough signal to get the silicon diode to work at all (although a much longer wire did provide a very small signal). A simple listening test and measurement of the AF prodiced by each diode showed that their was very little difference between the various germanium diodes. The MOSFET simply wired as a diode (gate wired directly the drain) was a poor detector - better than a silicon but not nearly so good as any of the germanium diodes. Finally the EPAD MOSFET wired as a syncronious detector worked really very well apparently producing ca. 10 times the signal (in other words it has much less loss than a germanium diode) for this particular set up and station than the germanium diodes.

As the syncronious arrangement was so succesful it got me thinking about this circuit. As the gate goes straight in on the main resonant circuit where the voltages, in a good Q circuit, might well be relatively high I thought I would try a standard FET (2N3819) and MOSFET (VN10KM) instead - and they worked! (they are a lot cheaper than an EPAD MOSFET chip and easier to get). Some FETS have a larger input capacitance so it helped to put a 10pf capacitor in series to the gate (a 10M ohm from gate to earth might reduce 50Hz pick up).

They were not as good as the EPAD MOSFET (although I only tried a single device) but the arrangment was still better than a simple 'crystal set' diode. For weak signals the EPAD MOSFET should perform better as it is able to keep going while the FET and the standard MOSFET will start to trail off as the resonant circuit voltage drops below the relatively higher gate voltage required for these devices. However it was interesting to see that it seems to be possible to get a more efficent set-up for a 'crystal' radio using a standard FET or MOSFET than with the standard germanium diode detector.

By the way I recently re-tried the 1N4148 diode but wired a 1.5V AA battery with a 100k resistor across it to forward bias it - and it worked ! It seemed to work better than a OA91!

SHORT WAVE 'CRYSTAL SET RADIO (bottom in photo)
The bottom circuit shown in the photo above is my own short wave version of the 'crystal set' radio but based on the references above. It's the same circuit as the MW radio but has a reduction drive tuning capacitor and different coil. I used about 7 turns (ca. 4cm diameter) for the resonant circuit and a couple of turns of insulated wire (white) for the coupling coil. I guess the radio tunes from about 3-8 MHz - I havent tried a sig gen on it yet.

I made this in the summer months (not the best time for listening on this part of the spectrum). Using an Earth and few meters of wire for antenna I heard 2 or 3 stations from around the world during the day while at night there were at least 7 including the voice of Russia world service, Radio Thailand as well as German and French speaking stations.

BACK TO G1EXG RADIO PAGE

THE CREATIVE SCIENCE CENTRE

Dr Jonathan Hare, The University of Sussex
Brighton, East Sussex. BN1 9QJ

home | diary | whats on | CSC summary | latest news

Design Issues in Radio Frequency Energy Harvesting System

Chomora Mikeka and Hiroyuki Arai
Yokohama National University
Japan

1. Introduction

Emerging self powered systems challenge and dictate the direction of research in energy harvesting (EH). State of the art in energy harvesting is being applied in various fields using different single energy sources or a combination of two or more sources. In certain applications like smart packaging, radio frequency (RF) is the preferred method to power the electronics while for smart building applications, the main type of energy source used is solar, with vibration & thermal being used increasingly. The main differences in these power sources is the power density; for example RF (0.01 ~ 0.1 $\mu W/cm^2$), Vibration (4 ~ 100 $\mu W/cm^2$), Photovoltaic (10 $\mu W/cm^2$ ~ 10mW/cm^2) and Thermal (20 $\mu W/cm^2$ ~ 10mW/cm^2). Obviously RF energy though principally abundant, is the most limited source on account of the incident power density metric, except when near the base stations. Therefore, in general, RF harvesting circuits must be designed to operate at the most optimal efficiencies.

This Chapter focuses on RF energy harvesting (EH) and discusses the techniques to optimize the conversion efficiency of the RF energy harvesting circuit under stringent conditions like arbitrary polarization, ultra low power (micro or nanopower) incidences and varying incident power densities. Harvested power management and application scenarios are also presented in this Chapter. Most of the design examples described are taken from the authors' recent publications.

The Chapter is organised as follows. Section 2.1 is the introduction on RF energy sources. Section 2.2 presents the antenna design for RF EH in the cellular band as well as DTV band. The key issue in RF energy harvesting is the RF-to-DC conversion efficiency and is discussed in Section 2.3, whereas Section 2.4 and 2.5 present the design of DTV and cellular energy harvesting rectifiers, respectively. The management of micropower levels of harvested energy is explained in Section 2.6. Performance analysis of the complete RF EH system is presented in Section 3.0. Finally, conclusions are drawn in Section 4.0.

1.1 RF energy sources

These include FM radio, Analogue TV (ATV), Digital TV (DTV), Cellular and Wi-Fi. We will present a survey of the measured E-field intensity (V/m) for some of these RF sources as shown in Table 1, [1]-[2]. Additionally, measured RF spectrums for DTV and Cellular signals are presented as shown in Fig. 1 to show on the potential for energy-harvesting in

these frequency bands. In general, many published papers on RF-to-DC conversion, have presented circuits capable of converting input or incident power as low as -20dBm. This means that, if an RF survey or scan finds signals in space, with power spectrum levels around -20dBm, then, it is potentially viable to harvest such signal power. In Fig. 1 (left side), the spectrum level is well above -20dBm and hence, a higher potential for energy harvesting. In Fig. 1 (right side), while the spectrum level is below -20dBm, what we observe is that the level increases with decrease in the distance toward the base station (BTS). Using free space propagation equation with this data, it was calculated that at a distance 1.4 m from the BTS, the spectrum level could measure 0dBm. An example calculation and plot for the estimated received power level, assuming 0dBi transmitter (BTS) and receiver antenna gains and free space propagation loss (FSPL) for FM and DTV is presented in Section 2.1.1. For the example estimation in Section 2.1.1, we select FM and DTV because they measured with a higher level than cellular and Wi-Fi for example.

Source	V/m	dBm	Reference	
FM radio	0.15~3		Asami et al.	
Analogue TV	0.3~2			
Digital TV	0.2~2.4	-40~0.0	Asami et al.	Arai et al.
Cellular		-65~0.0	Mikeka and Arai	
Wi-Fi		$\cong$ -30		

Table 1. RF energy sources, measured data.

In Table 1, FM radio has the highest E-field intensity implying the highest potential for energy harvesting. However, due to the requirements for a large antenna size and the challenges for simulations and measurements at the FM frequency i.e. 100 MHz or less (See Section 2.2.3, example FM antenna at 80 MHz), this Chapter will focus on DTV (470~770 MHz band) and Cellular (2100 MHz band) energy harvesting.
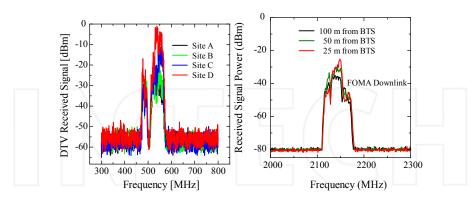
Fig. 1. DTV signal spectrum measured in Tokyo City (left side graph) and Cellular signal spectrum measured in Yokohama City (right side graph).

The received DTV signal power is high and also wide band, presenting high potential for increased energy harvesting unlike in cellular signals. We demonstrated in [2] that the total RF-to-DC converted power is roughly the integral over the DTV band (1), and is significantly larger than in the case of narrow band cellular energy harvesting.

$$P_{DC(DTV)} = \alpha \int_{470}^{770} \delta P_{DC}(f) df \; , \tag{1}$$

where α is the attenuation factor on the rectifying antenna's RF-to-DC conversion efficiency due to multiple incident signal excitation. δPDC is the small converted DC power from each of the single DTV signals in the 470 MHz to 770 MHz band.

In detail, we derive (1) from fundamentals as follows.

The incident power density on the rectifying antenna (rectenna), $S(\theta, \phi, f, t)$, is a function of incident angles, and can vary over the DTV spectrum and in time. The effective area of the antenna, $A_{eff}(\theta, \phi, f)$, will be different at different frequencies, for different incident polarizations and incidence angles. The average RF power over a range of frequencies at any instant in time is given by:

$$P_{RF}(t) = \frac{1}{f_{high} - f_{low}} \int_{f_{low}}^{f_{high}} \int_{0}^{4\pi} S(\theta, \phi, f, t) A_{eff(\theta, \phi, f)} d\Omega df \tag{2}$$

The DC power for a single frequency (f_i) input RF power, is given by

$$P_{DC}(f_i) = P_{RF}(f_i, t).\eta \left(P_{RF}(f_i, t), \rho, Z_{DC} \right), \tag{3}$$

where η is the conversion efficiency, and depends on the impedance match $\rho(P_{RF}, f)$ between the antenna and the rectifier circuit, as well as the DC load impedance. The reflection coefficient in turn is a nonlinear function of power and frequency.

The estimated conversion efficiency is calculated by P_{RF}/P_{DC}. This process should be done at each frequency in the range of interest. However, DC powers obtained in that way cannot be simply added in order to find multi-frequency efficiency, since the process is nonlinear. Thus, if simultaneous multi-frequency or broadband operation like in DTV band is required, the above characterization needs to be performed with the actual incident power levels and spectral power density. In this Chapter, we shall demonstrate DTV spectrum power harvest, given a rectenna than has been characterised in house at each single frequency in the DTV band.

1.1.1 An example calculation and plot for the estimated received power level

In this example we consider Tokyo's DTV and FM base stations (BS) as the RF sources. Both DTV and FM BS transmitter power (P_t) equals 10 kW (70dBm). The antenna gains are assumed 0dBi in both cases but also at the points of reception for easiness of calculation but with implications as follows. Assuming 0dBi antenna at each reception point, demands that we specify the frequency of the transmitted signal. For this reason we specify DTV signal frequency to be equal to 550 MHz while the FM signal frequency equals 80 MHz (Tokyo FM).

The received power, P_r is calculated using the simplest form of Friis transmission equation given by

$$P_r = P_t + G_t + G_r + FSPL \,, \tag{4}$$

where P_t = 70dBm, $G_t = G_r$ = 0dBi. G_r is the receiving antenna gain while FSPL is the free-space path loss given by

$$FSPL(dB) = 20\log(d) + 20\log(f) + 32.45 \,, \tag{5}$$

where d is in (km) and f is in (MHz). The plot for the received power as a function of distance from the DTV and FM base stations is shown in Fig. 2.
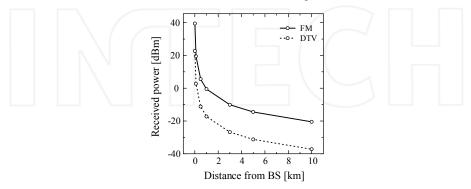
Fig. 2. DTV and FM received signal power level against distance.

With respect to Fig. 2, FM registers higher received power level than DTV at every reception point due to its lower transmit frequency and hence lower free-space path loss. For example at 1 km distance, FM received power level is -0.51dBm while for DTV, the received power is -17.26dBm. The important thing however, is that the received power level is frequency independent. It means that P_t is the transmitter power and the received power level at the position of distance d is $\dfrac{P_t}{4\pi d^2}$. However, if we assume 0dBi antenna at each reception point as in the above example, the power level is different because the antenna size of 0dBi is frequency dependent. As a result, high transmit power level is favorable for RF energy harvesting. Also near the base station is favorable.

1.2 Antenna design for the proposed RF energy harvesting (EH) system

It is well known that RF EH system requires the use of antenna as an efficient RF signal power receiving circuit, connected to an efficient rectifier for RF-to-DC power conversion. Depending on whether we want to harvest from cellular or DTV signals, the antenna design requirements are different. We will discuss the specific designs in the following sub sections.

1.2.1 Cellular energy harvesting antenna design

We propose a circular microstrip patch antenna (CMPA) for easy integration with the proposed rectifier (Section 2.5.1). However, the use of circular microstrip patch antennas (CMPA) is often challenged by the need for impedance matching, circular polarization (CP) and higher order harmonic suppression.

To address the above concerns, we create notches on the circular microstrip patch antenna. In our approach, we use only two, thin, fully parameterized triangular notches to achieve higher order harmonic suppression, impedance matching and circular polarization, all at once. This is the novelty in our proposed antenna. Our proposed CMPA is shown in Fig. 3. We study the behaviour of CMPA surface current vectors when notches (triangles ABC) are created on the structure at $\alpha = 45^0$ and $\alpha = 225^0$.
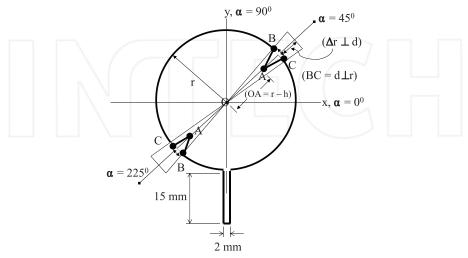
Fig. 3. Cellular energy harvesting antenna structure.

Notch parameters *d* and *h* in Fig. 3 were investigated by calculation using CST microwave Studio.

Without notches, the CMPA's input is not matched at f_c= 2.15 GHz as shown in Fig. 4 (left side). However, with notches, matching is achieved. The parameter combination *d* = 7 and *h* = 6 offers a matched and widest band input response and hence we adopt it for cellular energy harvesting applications.
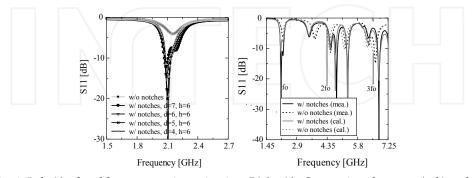
Fig. 4. *Left side: d* and *h* parameter investigation. *Right side:* Comparison between (cal.) and (mea.) S_{11} at f_0 = 2.175 GHz, $2f_0$ = 4.35 GHz and $3f_0$ = 6.53 GHz. The adopted notch parameters are *d*=7 mm while *h*= 6 mm.

The comparison between calculated and measured S_{11} is shown in Fig. 4 (right side). The 2nd and 3rd harmonics are suppressed as required by design. The comparison between calculated and measured radiation patterns is shown in Fig. 5, where $E_\vartheta \cong E_\varphi$ due to the 45^0 tilted surface current vector. Ordinarily, without notches, the surface current vector is parallel to the microstrip feeder axis. In conclusion, our proposed CMPA is sufficiently able to suppress higher order harmonics while simultaneously radiating a circularly polarized (CP) wave. The CP is required to efficiently receive the arbitrary polarization of the incident cellular signals at the rectenna.
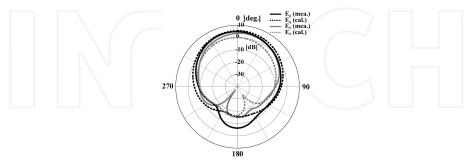
Fig. 5. Cellular energy harvesting antenna pattern at f_c= 2.15 GHz.

1.2.2 DTV energy harvesting antenna design

Unlike in the cellular energy harvesting antenna, the DTV energy harvesting antenna must be wideband (covering 470 MHz to 770 MHz), horizontally polarized and omni-directional.

The proposed antenna is typically a square patch (57 mm x 76 mm) with a partial ground plane (9 mm x 100 mm). The patch is indirectly fed by a strip line (9 mm x 3 mm). The proposed antenna geometry is shown in Fig. 6. The partial ground plane is used to achieve omni-directivity and a certain level of wide bandwidth. To tune the impedance of this antenna as well as to adjust the bandwidth within the target band, a "throttle" with stepped or graded structures is used between the microstrip feed line and the square patch, as shown in Fig. 6 (left side).
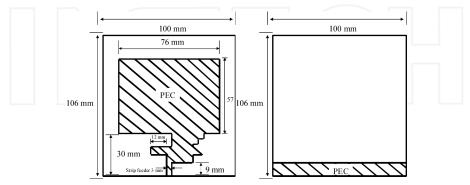
Fig. 6. Proposed DTV antenna geometry. *Left side:* Front view. *Right side:* Back view. The antenna is printed on FR4 substrate; t = 1.6 mm, ε_r = 4.4.

The input response for the proposed antenna is shown in Fig. 7 (left side). The omni directivity is confirmed by measurement at 500 MHz, 503 MHz, and 570 MHz as shown in Fig. 7 (right side). The radiation patterns shown in Fig. 7 are for the xz plane, which happens to be the vertical polarization for the antenna. DTV signals are horizontally polarized and therefore, when using this antenna, the orientation must be in such a way as to efficiently receive the DTV signal. Simply a 90 degree rotation of the antenna along the z axis achieves this requirement.
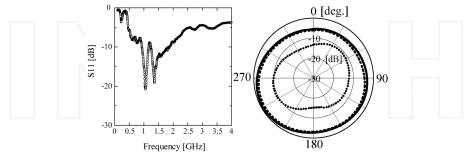
Fig. 7. Proposed DTV antenna performance. *Left side:* The antenna's measured input performance. *Right side:* The omni directivity in the vertical plane is confirmed at 500 MHz, 503 MHz, and 570 MHz. The outermost, black solid and dotted line patterns represent 503 MHz and 500 MHz directivity, respectively. The innermost dotted line pattern is the directivity at 570 MHz.

1.2.3 Example design for an 80 MHz FM half-wave dipole antenna

A half-wave dipole is the simplest practical antenna designed for picking up electromagnetic radiation signals, see Fig. 8 (courtesy of Highfields Amateur Radio Club). Calculating the optimal antenna length to pick up a certain frequency signal is fairly straightforward because antenna physics demand that the total length of wire used in the antenna be equal to one wavelength of the type of electromagnetic radiation it will be picking up. This means that the total length of the antenna should be equal to half the desired wavelength. By converting the 80 MHz frequency into a wavelength, you can thus obtain your antenna length as 1.875m by using the magic equation, $\lambda = \dfrac{c}{f}$. However, the actual length is typically about 95% of a half wavelength in free space, hence a half-wave dipole for this frequency should be 1.788m long, which would make each leg of the dipole 0.894m in length.

1.3 RF-to-DC conversion efficiency improvement techniques

A Schottky diode circuit connected to an antenna is used for RF-to-DC power conversion. To convert more of the antenna surface incident RF power to DC power, high RF-to-DC conversion efficiency is required of the rectifying circuit. Many authors have shown that the efficiency depends on several factors like Schottky diode type, harmonics suppression capability, load resistance selection, and the capability to handle arbitrary polarized incident waves. What is missing in most of these published works is the efficiency optimization for
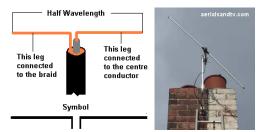
Fig. 8. Half-wave dipole. *Left side:* Antenna structure. *Right side:* Typical deployment.

ultra low power incident waves and the explanation of the physical phenomena behind most of the recommended efficiency optimization approaches.

This Chapter will show for example that a Schottky diode that delivers the highest efficiency at 0dBm incidence may not necessarily deliver the highest efficiency at lower power incidence e.g. –20dBm. We will therefore classify which diodes perform better at given power incidences; of course, this will also be compared to the diode manufacturers' application notes. Simulations in Agilent's ADS using SPICE and equivalent circuit models will compare the performance of few selected Schottky diodes namely; HSMS-2820, HSMS-2850, HSMS-2860, HSC-276A, and SMS7630. Moreover, the effect of the Schottky diode's junction capacitance (C_j) and junction bias resistance (R_j) on the conversion efficiency will be shown from which, special techniques for Schottky diode harmonic suppression and rectifying circuit loading for maximum efficiency point tracking will be presented.

1.3.1 The schottky diode

The classical *pn* junction diode commonly used at low frequencies has a relatively large junction capacitance that makes it unsuitable for high frequency application [3]. The Schottky barrier diode, however, relies on a semiconductor-metal junction that results in a much lower junction capacitance. This makes Schottky diodes suitable for higher frequency conversion applications like rectification (RF-to-DC conversion) [3]. We will demonstrate the effects of junction capacitance and resistance in the following sub section.

1.3.2 The effect of Schottky diode's C_j and R_j on the conversion efficiency

We have studied Schottky diode's C_j and R_j and published our results in [4]. In this work, we designed a rectifying antenna tuned for use at 2 GHz. The circuit proposed in [4] is a voltage doubler by configuration, but we replaced the amplitude detection diode (series diode) with its equivalent circuit adapted from [5]. The results of this investigation show that variation of C_j shifts the tuned frequency position and also introduces a mismatch in the resonant frequency, see Fig. 9 (left side graph). Therefore for this circuit at 2 GHz, we recommend using a Schottky diode having C_j = 0.2pF. In general, a smaller value of C_j is desirable at higher frequencies. Similarly, for R_j investigation, a smaller value is desirable for better matching at 2 GHz for example. If the R_j is increased towards 10kΩ, there is a mismatch in the resonant frequency but no shift in the frequency, see Fig. 9 (right side graph). Another approach to the study of Schottky diodes for higher frequency and efficiency rectenna design is presented in [6].
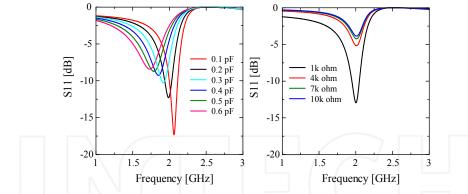
Fig. 9. Schottky diode's C_j effect (Left) and R_j effect (Right) on 2 GHz rectenna's input response.

1.4 Rectifying circuit for DTV energy harvesting

In the design of a DTV energy harvesting circuit, several basic design considerations must be paid attention to. First is the antenna; it must be wideband (covering 470 MHz to 770 MHz), horizontally polarized and omni-directional. Secondly is the rectifier; it must also be wideband, and optimized for RF-to-DC conversion for incident signal power at least -40dBm. Until recently, very few authors have published on DTV energy harvesting circuit. For the few publications, the antenna could not meet all those three requirements and a discussion on the performance of the harvesting circuit for ultra low power incidences has been neglected. In this Chapter we will present such a rectenna with conversion efficiencies above 0.4% at -40dBm, above 18.2% at -20dBm and over 50% at -5dBm signal power incidence. We will closely compare simulated and measured performance of the rectenna and discuss any observed disparities.

Agilent's ADS will be used to simulate the nonlinear behaviour of the rectifying circuit based on harmonic balance tuning methods. To simulate the multiple incident waves, a multi-tone excitation in the DTV band will be invoked. The wideband input characteristic will be achieved by the input matching inductors and capacitors.

The generic version of our proposed DTV energy harvesting circuit is shown below in Fig. 10. The implementation, however, is in two phases or scenarios as follows. First, we investigate the class called "ultra low power" DTV band rectenna. Secondly, we introduce the "medium power" DTV band rectenna.
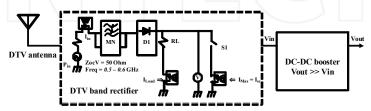
Fig. 10. Generic version of our proposed DTV energy harvesting circuit.

1.4.1 Ultra low power DTV rectenna

We define an ultra low power rectenna as one impinged by RF power incidence in the range between – 40dBm and -15dBm. Below in Fig. 11 is the circuit we designed; optimized for -20dBm input. The matching network is complex so as to achieve a wide band input characteristic. The fabricated circuit was well matched for the frequency range between 470 MHz and 600 MHz. More details about the circuit design can be found in [7].
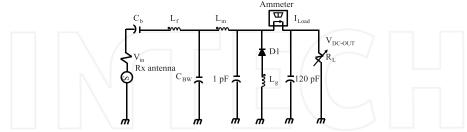
Fig. 11. Ultra low power DTV band rectenna circuit. SMS7630 Schottky diode by SKYWORKS offered the best performance.

The RF-to-DC conversion efficiency for this circuit is shown in Fig. 12 where at input power equal to -40dBm, efficiency is at least 0.4% and rectified voltage equals 1mV; at -20dBm, we have at least 18.2% by measurement and a rectified voltage of 61.7mV. The level of rectified voltage is too low and disqualifies this circuit for purposes of charging capacitors or batteries to accumulate such micropower over time. Instead, boosting the low voltage to usable levels is the option available and we shall discuss this at a later stage, (in Section 2.6).
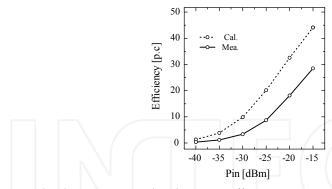
Fig. 12. Ultra low power DTV band rectenna efficiency.

1.4.2 Medium power DTV rectenna

We define a medium power rectenna as one impinged by RF power incidence in the range between – 5dBm and 0dBm. Below in Fig. 13 is the circuit we designed, optimized for -5dBm input. The matching network is simpler than as shown in section 2.4.1 since we require a narrow band around 550 MHz, with received peak power spectrum levels at least -5dBm. The circuit in Fig. 13 is a modification of Greinacher's doubler rectifier. In the circuit, C_b equals 1 pF and is used to block DC current against flowing towards the source. The shunt

capacitance, C_{BW} equals 3300 pF and is used to set the input bandwidth. The grounding inductance, L_g equals 56nH (optimal) and is used to improve the RF-to-DC conversion efficiency by cancelling the Schottky diodes (D_b and D_D) capacitive influence; thereby minimizing the harmonic levels (harmonic suppression). We used HSMS2850 diodes in these circuits for their better performance at this level of incident power.
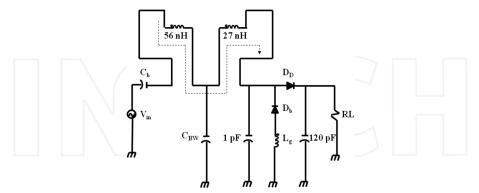
Fig. 13. Medium power DTV rectenna circuit. HSMS 2850 or 2820 from Hewlett-Packard offered the best performance.

The RF-to-DC conversion efficiency for this circuit is shown in Fig. 14 where at input power equal to -5dBm, we achieve at least 50% conversion efficiency by measurement, equivalent to 1.2 V DC rectified at 8.2kΩ optimal load. If we change the load to 47kΩ, over 2 V DC is rectified. This rectenna circuit is ideal for powering small sensors that run on 1.5 V or 2.2 V and draw around 6μA nominal current. If we need to power sensors demanding more power, say at least 2.2 V and 0.3mA to 1.47mA current consumption, we have to accumulate the power in a capacitor over time.
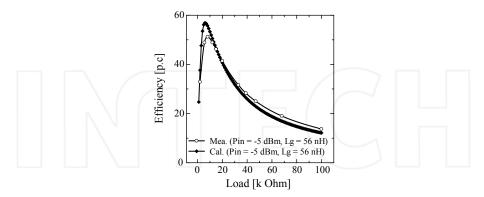
Fig. 14. Medium power DTV band rectenna efficiency.

1.4.3 DTV energy harvesting scenario and application demo

Using the medium power DTV band rectenna, connected to a gold capacitor as an accumulator, energy harvesting was initiated as shown in Fig. 15. Details about the gold

capacitor, which include its charge function, backup time and leakage losses are presented in [8]. For the scenario shown in Fig. 15, the accumulated voltage by measurement i.e. capacitor charge function follows the path;

$$V_{acc} = 0.5388\ln(t) + 1.4681 \tag{6}$$

where V_{acc} is the accumulated voltage in volts and t the time in hours. It takes 4.5 hours to accumulate 2.25 V, given a rectified charging voltage and current of 2.4 V and 51μA, respectively, supplied by the DTV band rectenna instantaneously.

With this rectenna, it was possible to power up many different kinds of sensors. Sensors with ultra low power consumption were powered directly, without need to accumulate the power in a capacitor, as shown in Fig. 16.

Fig. 15. DTV energy harvesting in a park at some line of sight from the base station.
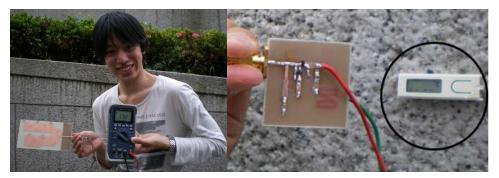
Fig. 16. Directly powering a thermometer mounted on a car park wall (right picture). The maximum instant voltage rectification on record equals 3.7 V (left picture).

1.5 Rectifying circuit for cellular energy harvesting

Unlike in the DTV energy harvesting circuit, for cellular energy harvesting, the antenna must be narrowband (50 MHz bandwidth is acceptable), and circularly polarized even

though cellular signals are vertically polarized. The circular polarization is desired to maximize the RF-to-DC conversion efficiency of the arbitrary polarization incident signals in the multipath environment. Similarly, the rectifier must be narrowband, and optimized for RF-to-DC conversion over a wide range of incident signal power.

Thinking about the potential applications for cellular energy harvesting is useful. Other authors have reported on powering a scientific calculator or a temperature sensor from GSM energy harvesting. In this Chapter we will present a special application for energy harvesting in the vicinity of the W-CDMA cellular base station and analyze the system performance by calculation from experimental data. A cellular energy harvesting circuit optimized for over 50% RF-to-DC conversion efficiency given approximately 0dBm incidence will be presented.

1.5.1 Cellular band rectenna

Below in Fig. 17 is the circuit we designed, optimized for 0dBm input. Simple input matching network is ideal since we require a narrow band response around 2.1 GHz. The optimum value for L_g equals 5.6nH, where L_g is used to improve the RF-to-DC conversion efficiency as earlier discussed. HSMS2850 diode was used.
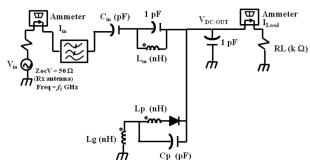
Fig. 17. Shunt rectifier configuration for the cellular band. The matching elements L_m = 3.2nH, while C_{in} =2.5pF. The load resistance is fixed at R_L = 2.1kΩ.
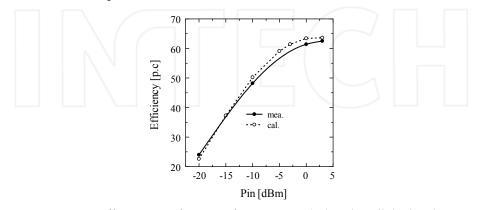
Fig. 18. Conversion efficiency as a function of input power (P_{in}) in the cellular band.

The RF-to-DC conversion efficiency for this circuit is shown in Fig. 18 where at input power equal to 0dBm, we achieve at least 60% conversion efficiency by measurement, given a 2.1kΩ optimal load. This rectenna circuit is ideal for powering small sensors that run on 1.5 V or 2.2 V and 6µA nominal current consumption. If we need to power sensors demanding more power, say at least 2.2 V and 0.3mA to 1.47mA, we have to accumulate the power in a capacitor over time as discussed in section 2.4.3 above.

1.5.2 Cellular energy harvesting application example

Environmental power generation in the neighbourhood of a cellular base station to power a temperature sensor is proposed as shown in Fig. 19 below. Electric field strength measurements in the base station neighbourhood have demonstrated the potential for environmental power generation, and the proposed temperature sensor system is designed based on these values. The rectenna described in Section 2.5.1 is used as the RF-to-DC rectifying circuit with the notched circular microstrip patch antenna (CMPA) proposed in Section 2.2.1. RF-to-DC conversion efficiency equal to 53.8% is obtained by measurement. The temperature sensor made for trial purposes clarifies the capability for temperature data wireless transmission for 20 seconds per every four hours in the base station neighbourhood.
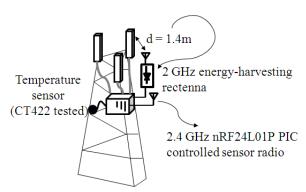
Fig. 19. Application example in the vicinity of the cellular base station.

1.6 Micropower energy harvesting management

A rectifying antenna circuit for -40dBm incident power harvesting generates 1mV at 2kΩ load, given 0.4% efficiency as presented in Section 2.4.1. At -20dBm incidence and at least 18.2% efficiency, 61.7mV is generated given a 2kΩ load [7]. The generated DC power in both of these two cases is in the µW range, hence the micropower definition. To manage such micropower, power accumulation or energy storage is required. Storage devices may either be a gold capacitor, super capacitor, thin film battery or the next generation flexible paper batteries. These storage devices have specific or standard maximum voltage and trickle charging current minimum requirements. Typically, gold capacitors have voltage ratings like 2.7 V, 5.5 V for 100 µA, 10mA or 100mA maximum discharge current. On the other hand, standard ratings for batteries are 1.8 V, 2 V, 3.3 V and 4.1 V. Therefore, to directly charge any of these storage devices from 1mV, or 61.7mV DC is impractical.

Published works have demonstrated the need for a DC-to-DC boost converter placed between the rectifying antenna circuit (rectenna) and the storage device. Recent efforts have demonstrated that a 40mV rectenna output DC voltage could be boosted to 4.1 V to trickle charge some battery. A Coilcraft transformer with turns ratio ($N_s : N_p$) equal to 100 was used in the boost converter circuit. An IC chip leading manufacturer (Linear Technology Corp., LT Journal, 2010) has released a linear DC-to-DC boost regulator IC chip capable of boosting an input DC voltage as low as 20 mV and supplying a number of possible outputs, specifically suited for energy harvesting applications. While this IC is a great milestone, readers and researchers need to understand the techniques to achieve such ICs and also the limitations that apply. In the following sub section, we will describe the methods toward designing a DC-DC boost converter, suitable for micropower RF energy harvesting.

In the design, we will attempt to clarify the parameters that affect the DC-DC conversion efficiency. For this design, Envelope simulation in Agilents's ADS is used. This simulation technique is the most efficient for the integrated rectenna and DC-DC boost converter circuits.

1.6.1 DC-DC boost converter design theory and operation

The DC-DC boost converter design theory and actual implementation are presented in this section. The inequality $V_{in} \ll V_{out}$ defines the boost operation. In this Chapter, our boost converter concept is illustrated in Fig. 20. A small voltage, V_{in} is presented at the input of the boost converter inductive pump which as a result, generates some output voltage, V_{out}. The output voltage is feedback to provide power for the oscillator. The oscillator generates a square wave, F_{OSC} that is used for gate signalling at the N-MOSFET switch.
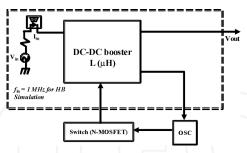
Fig. 20. Boost converter concept.

The drain signal of the N-MOSFET is used as the switch node voltage, V_{sn} at the anode of the diode inside the boost converter circuit block. From the concept presented in Fig. 20, the actual implemented circuit is shown in Fig. 21. The circuit was designed in Agilent's ADS and fabricated for investigation by measurement.

The circuit in Fig. 21 is proposed for investigation. Since a DC-DC boost converter is supposed to connect to the rectenna's output, it therefore, becomes the load to the rectenna circuit. This condition demands that the input impedance of the boost converter circuit emulates the known optimum load of the rectenna circuit. This has the benefit of ensuring

maximum power transfer and hence higher overall conversion efficiency from the rectenna input (RF power) to the boost converter output (DC power). In this investigation, as shown in [7], the optimum load for the rectenna is around 2kΩ. In general, emulation resistance R_{em} is given by
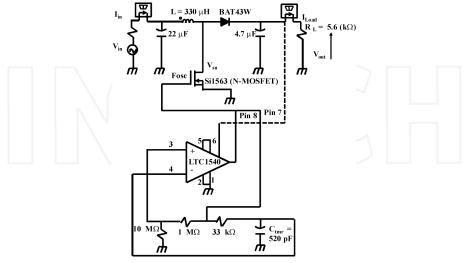
Fig. 21. The proposed boost converter circuit diagram. Designed in Agilent's ADS and fabricated for investigation by measurement.

$$R_{em} = \frac{2LT}{t_1^2 k}\left(\frac{M-1}{M}\right) \tag{7}$$

where L is the inductance equal to 330µH as shown in Fig. 20, $M = \frac{V_{out}}{V_{in}}$, T is the period of F_{OSC}, t_1 is the switch"ON" time for the N-MOSFET, and k is a constant that according to [3] is a low frequency pulse duty cycle if the boost converter is run in a pulsed mode and typically, k may assume values like 0.06 or 0.0483. With reference to (7), we select L as the key parameter for higher conversion efficiency while V_{in} = 0.4 V DC is selected as the lowest start up voltage to achieve oscillations and boost operation. Computing the DC-DC boost conversion efficiency against different values of L, we have results as shown in Fig. 22.

From the results above, L = 100µH is the optimum boost inductance that ensures at least 16.5% DC-DC conversion efficiency, given R_L = 5.6kΩ.

Now having selected the optimum boost inductance given some load resistance, the emulation resistance shown in Fig. 23 is evaluated from the ratio of voltage versus current at the boost converter circuit's input.

The results show a constant resistance value against varying inductance. In general, we can say that this boost converter circuit has a constant low input impedance around 82.5Ω. This impedance is too small to match with the optimum rectenna load at 2kΩ. This directly affects the overall RF-to-DC conversion efficiency.

The results show a constant resistance value against varying inductance. In general, we can say that this boost converter circuit has a constant low input impedance around 82.5Ω. This impedance is too small to match with the optimum rectenna load at 2kΩ. This directly affects the overall RF-to-DC conversion efficiency.
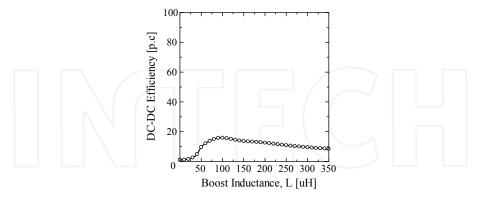
Fig. 22. Boost inductance variation with DC-DC conversion efficiency for a 5.6 kΩ load.
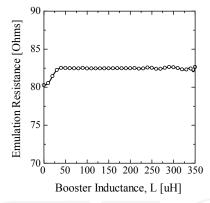
Fig. 23. Boost converter's input impedance: the emulation resistance.

Another factor, which affects the overall conversion efficiency is the power lost in the oscillator circuit. Unlike the circuit proposed in [9], which uses two oscillators; a low frequency (LF) and high frequency (HF) oscillator; in Fig. 21, we have attempted to use a single oscillator based on the LTC1540 comparator, externally biased as an astable multivibrator.

The power loss in this oscillator is the difference in the DC power measured at Pin 7 (supply) to the power measured at pin 8 (output). We term this loss, L_{osc}; converted to heat or sinks through the 10MΩ load. A comparison of the oscillator power loss to the power available at the boost converter output is shown in Fig. 24.

Looking at Fig. 24; we notice that the power loss depends on whether the oscillator output is high or low. The low loss corresponds to the quiescent period where the power lost is

almost negligible. However, during the active state, the lost power (power consumed by the oscillator) nearly approaches the DC power available at the boost converter output. This results in low operational efficiency.
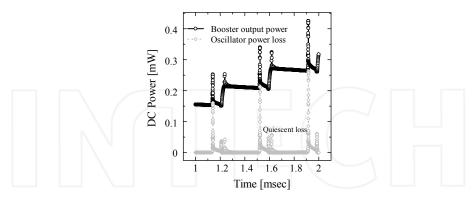
Fig. 24. The power loss in the oscillator.

To confirm whether or not the circuit of Fig. 21 works well, we did some measurements and compared them with the calculated results. Unlike in calculation (simulation), during measurement, L = 330µH was used due to availability. All the other component values remain the same both in calculation and measurement. In Fig. 25 (left side graph) and (right side graph), we see in general that the input voltage is boosted and also that the patterns of F_{osc} and V_{sn} are comparable both by simulation and measurement. To control the duty cycle of the oscillator output (F_{osc}), and the level of ripples in the boost converter output voltage (V_{out}), we change the value of the timing capacitance, C_{tmr} in the circuit of Fig. 21. Simulations in Fig. 25 (left side graph) show that C_{tmr} = 520pF realizes a better performance i.e. nearly constant V_{out} level (very low ripple).
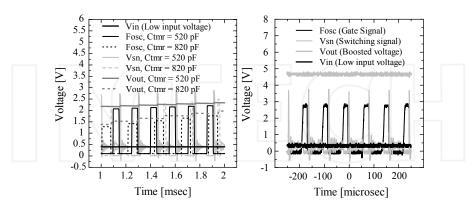
Fig. 25. Voltage characteristics of the developed boost converter circuit. The left side graph represents simulation while the right side graph is for measurements.

Generally, we observe that with this kind of boost converter circuit topology, it is difficult to start up for voltages as low as 61.7mV DC generated by the rectenna at -20dBm power

incidence and at least 18.2% rectenna RF-to-DC conversion efficiency. Self starting is the issue for this topology at very low voltages.

At least 11.3% DC-DC conversion efficiency was recorded by measurement and is comparable to the calculation in Fig. 22. During measurement it was clearly revealed that the boost converter efficiency does depend on the value of **L** and the duty cycle derived from t_1. To efficiently simulate the complete circuit, from the RF input to the DC output, envelope transient simulation (ENV) in Agilent's ADS was used. The (ENV) tool is much more computationally efficient than transient simulation (Tran). This simulation is appropriate for the boost converter circuit's resistor emulation task. Moreover, the boost converter's DC-DC conversion efficiency, and the overall RF-to-DC conversion efficiency can be calculated at once with a single envelope transient simulation.

In summary, though not capable to operate for voltages as low as 61.7mV DC, the proposed boost converter has by simulation and measurement demonstrated the capability to boost voltages as low as 400mV DC, sufficient for battery or capacitor recharging, assuming that the battery or the capacitor has some initial charge or energy enough to provide start-up to the boost converter circuit.

The limitations of our proposed boost converter circuit include; low efficiency, lack of self starting at ultra low input voltages, and unregulated output. To address these limitations, circuit optimization is required. Moreover, alternative approaches which employ a flyback transformer to replace the boost converter inductance must be investigated. A regulator circuit with Low Drop Out (LDO) is necessary to fix the boost converter output voltage commensurate with standard values like 2.2 V DC for example. For further reading, see [7]

2. Performance analysis of the complete RF energy harvesting sensor system

To demonstrate how one may analyze the performance of an RF energy harvesting system including its application, we extend the discussion of Section 2.5.2 to this Section. We propose a transmitter assembled as in Fig. 26 for temperature sensor wireless data transmission.
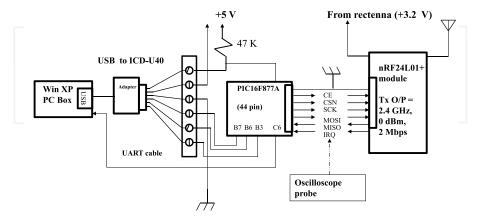
Fig. 26. The assembly and test platform for the proposed battery-free sensor transmitter.

The transmitter consists of one-chip microcomputer (MCU) PIC16F877A and wireless module nRF24L01P for the control, and MCU can be connected with an outside personal computer using ICD-U40 or RS232 cable. The wireless module operates in transmission and reception mode, and controls power supply on-off, transmitting power level, the receiving mode status, and transmission data rate via Serial Peripheral Interface (SPI). Figure 27 shows the operation flow when transmitting.
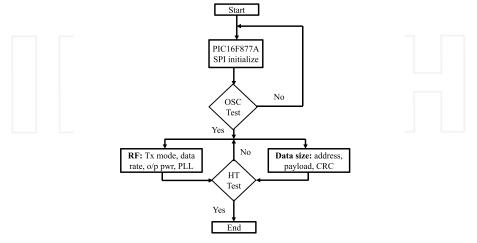
Fig. 27. Operation flow during transmission.

The experimental system composition is shown in Fig. 28 to transmit acquired data by the temperature sensor with WLAN at 2.4 GHz (ISM band). An ISM band sleeve antenna is used for the transmission. Using the cellular band rectenna shown and discussed in Section 2.5.1, at least 3.14 V is stored in the electric double layer capacitor over a period of four hours. To harvest a maximum usable power for the overall system, we charge the capacitor up to 5V. The operation voltage for the wireless module presented in Fig. 26 above is between 1.9V and 3.6V.

The signal was transmitted from the wireless module while a sleeve antenna, same like the one for transmission was used with the spectrum analyzer and the reception experiment was performed. Received signal level equal to -43.4dBm was obtained at a distance 3.5m between transmitter and reception point. The capacitor's stored voltage was used to supply the wireless module in the above-mentioned experiment. Successful transmission was possible for 5.5 minutes after which, the capacitor terminal voltage decreased from 3.16V to 1.47V, and the transmission ended. The sending and receiving distance of data can be estimated to be about 10m when the sensitivity of the receiver is assumed to be -60dBm, given 0dBm maximum transmit power.

Hereafter, the overall system examination is done by environmental power generation using the transmitted electric waves from the cellular phone base station, proposed based on the above-mentioned results. First of all, the power consumption shown in Fig. 29 is based on the fact that 120mW (5V, 24mA) is saved in the electric double layer capacitor by environmental power generation, achieved by calculation as discussed earlier.
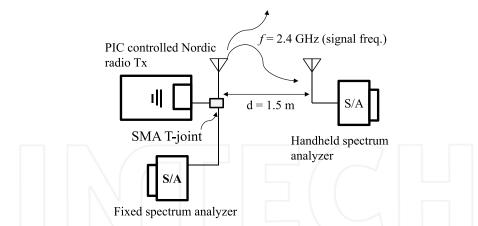
Fig. 28. Indoor measurement setup for received traffic from the sensor radio transmitter.
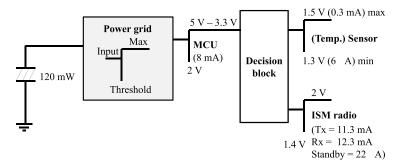
Fig. 29. Power management scheme for the cellular energy-harvesting sensor node.

The sensor data packet is transmitted wirelessly in ShockBurst mode for energy efficient communication. The data packet format includes a pre-amble (1 byte), address (3 bytes), and the payload i.e. temperature data (1 byte). The flag bit is disregarded for easiness, and cyclic redundancy check (CRC) is not used.

The operation of the proposed system is provisionally calculated. When the rectenna is set up in the place where power incidence of 0dBm is obtained in the base station neighbourhood (as depicted in Section 2.5.2), an initially discharged capacitor accumulates up to 3.3V by a rectenna with 53.8% conversion efficiency (presented in Section 2.5.1). At this point, it takes 1.5 minutes to start and to initialize a wireless module, and the voltage of the capacitor decreases to 2V. This trial calculation method depends on the capacitor's back up time discussed in [8]. After this, when the wireless module is assumed to be in sleep mode, the capacitor is charged by a 0.28mA charging current for four hours whereby the capacitor's stored voltage increases up to 5V. The power consumption in the sleep mode or standby is 33μW (1.5V, 22μA).

When the wireless module starts, after data transmission and the confirmation signal is sent, the voltage of the capacitor decreases by 0.6V, and consumes the electric power of 7.4mW.

The voltage of the capacitor decreases to 2V when 3.2mW is consumed to the acquisition of the sensor data, and the operation time of MCU is assumed to be one minute to the data storage in the wireless module etc. As for the capacitor voltage, when the wireless module continuously transmits data for 20 seconds, it decreases from 2V to 1.4V and even the following operation saves the electric power. Therefore, a temperature sensing system capable of transmitting wireless data in every four hours becomes feasible by environmental power generation from the cellular phone base station if we consider intermittent operation by sleep mode.

3. Conclusion
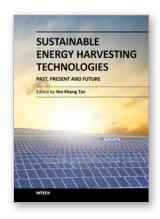
This Chapter has given an overview of the present energy harvesting sources, but the focus has stayed on RF energy sources and future directions for research. Design issues in RF energy harvesting have been discussed, which include low conversion efficiency and sometimes low rectified power. Solutions have been suggested by calculation and validated by measurement where possible, while highlighting the limitations of the proposed solutions. Potential applications for both DTV and cellular RF energy harvesting have been proposed and demonstrated with simple examples. A discussion is also presented on the typical performance analysis for the proposed RF energy harvesting system with sensor application.

4. Acknowledgment

5. References

[1] Keisuke, T.; Kawahara, Y. & Asami, T. (2009). *RF Energy Intensity Survey in Tokyo..,* (c)2009 IEICE, B-20-3, Matsuyama-shi, Japan

[2] Mikeka, C.; Arai, H. (2011). *Dual-Band RF Energy-Harvesting Circuit for Range Enhancement in Passive Tags,* (c)2011 EuCAP, Rome, Italy

[3] Pozar, D. (2005). *Microwave Engineering,* Wiley, ISBN 978-0-471-44878-5, Amherst, MA, USA

[4] Mikeka, C.; Arai, H. (2010). Techniques for the Development of a Highly Efficient Rectenna for the Next Generation Batteryless System Applications, *IEICE Tech. Rep.,* pp. 101-106, Kyoto, Japan, March, 2010

[5] http://www.secomtel.com/UpFilesPDF/PDF/Agilent/PDF_DOCS/SKYDIODE/ 03_ SKYDI/HSMS2850.PDF (Last accessed on 13 July, 2011)

[6] McSpadden, J. et al., H. (1992). Theoretical and Experimental Investigation of a Rectenna Element for Microwave Power Transmission, *IEEE Trans., on Microwave Theory and Tech., Vol. 40, No. 12.,* pp. 2359-2366, Dec., 1992

[7] Mikeka, C.; Arai, H. ; Georgiadis A. ; and Collado A. (2011). DTV Band Micropower RF Energy-Harvesting Circuit Architecture and Performance Analysis, *RFID-TA Digest,* Sitges, Spain, Sept., 2011

[8] Mikeka, C.; Arai, H. (2009). Design of a Cellular Energy-Harvesting Radio, *Proc. 2 nd European Wireless Technology Conf.,* pp. 73-76, Rome, Italy, Sept., 2009

[9] Popovic Z., et al., (2008). Resistor Emulation Approach to Low-Power RF Energy Harvesting, *IEEE Trans. Power Electronics, Vol. 23, No. 3,* 2008

Sustainable Energy Harvesting Technologies - Past, Present and Future

Edited by Dr. Yen Kheng Tan

In the early 21st century, research and development of sustainable energy harvesting (EH) technologies have started. Since then, many EH technologies have evolved, advanced and even been successfully developed into hardware prototypes for sustaining the operational lifetime of low?power electronic devices like mobile gadgets, smart wireless sensor networks, etc. Energy harvesting is a technology that harvests freely available renewable energy from the ambient environment to recharge or put used energy back into the energy storage devices without the hassle of disrupting or even discontinuing the normal operation of the specific application. With the prior knowledge and experience developed over a decade ago, progress of sustainable EH technologies research is still intact and ongoing. EH technologies are starting to mature and strong synergies are formulating with dedicate application areas. To move forward, now would be a good time to setup a review and brainstorm session to evaluate the past, investigate and think through the present and understand and plan for the future sustainable energy harvesting technologies.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Chomora Mikeka and Hiroyuki Arai (2011). Design Issues in Radio Frequency Energy Harvesting System, Sustainable Energy Harvesting Technologies - Past, Present and Future, Dr. Yen Kheng Tan (Ed.), ISBN: 978-953-307-438-2, InTech, Available from: http://www.intechopen.com/books/sustainable-energy-harvesting-technologies-past-present-and-future/design-issues-in-radio-frequency-energy-harvesting-system

INTECH
open science | open minds

Multi-Service Highly Sensitive Rectifier for Enhanced RF Energy Scavenging

Negin Shariati, Wayne S. T. Rowe, James R. Scott & Kamran Ghorbani

School of Electrical and Computer Engineering, RMIT University, Melbourne, VIC 3001, Australia.

Due to the growing implications of energy costs and carbon footprints, the need to adopt inexpensive, green energy harvesting strategies are of paramount importance for the long-term conservation of the environment and the global economy. To address this, the feasibility of harvesting low power density ambient RF energy simultaneously from multiple sources is examined. A high efficiency multi-resonant rectifier is proposed, which operates at two frequency bands (478–496 and 852–869 MHz) and exhibits favorable impedance matching over a broad input power range (-40 to -10 dBm). Simulation and experimental results of input reflection coefficient and rectified output power are in excellent agreement, demonstrating the usefulness of this innovative low-power rectification technique. Measurement results indicate an effective efficiency of 54.3%, and an output DC voltage of 772.8 mV is achieved for a multi-tone input power of -10 dBm. Furthermore, the measured output DC power from harvesting RF energy from multiple services concurrently exhibits a 3.14 and 7.24 fold increase over single frequency rectification at 490 and 860 MHz respectively. Therefore, the proposed multi-service highly sensitive rectifier is a promising technique for providing a sustainable energy source for low power applications in urban environments.

AMBIENT energy harvesting is attracting widespread interest as it has the potential to provide a sustainable energy source for future growth and protection of the environment. Considerable research effort has been directed toward low-profile, low-power, energy efficient and self-sustainable devices aiming to harvest energy from inexhaustible sources such as solar energy, thermal, biomass, mechanical sources (e.g. wind, kinetic, vibration, and ocean waves) wastewater, and microwave energy. A thorough set of reviews is given in the literature[1–4]. Among these green energy sources, there has been a growing interest for radio frequency (RF) energy scavenging, as the availability of ambient RF energy has increased due to advancements in broadcasting and wireless communication systems. Furthermore, the development of wireless power transmission (WPT) technologies[5] that allow micro sensors[6], mobile electronic devices[7], wireless implantable neural interfaces[8] and far-field passive RFID (Radio-Frequency Identification) systems[9–11] to operate without batteries has triggered impetus for RF energy harvesting.

Efficient RF energy harvesting is a very challenging issue, as it deals with the very low RF power levels available in the environment. Furthermore, the scavengeable power level can vary unpredictably, depending on several factors such as the distance from the power source, the transmission media, the telecommunication traffic density and the antenna orientation. The majority of available literature on RF rectification has been dedicated to narrowband rectennas, which essentially operate at a single frequency and hence provide low DC output power[12,13]. Various topologies, such as voltage doublers or multipliers have been employed in order to increase the RF to DC conversion efficiency and the output DC voltage for specific applications[14–16]. However, from an ambient RF scavenging perspective, harvesting energy from various available frequencies could maximize power collection and hence increase the output DC power. Ultra-wideband and broadband rectenna arrays have been proposed as a potential solution[17,18]. However in some cases, simulation and experimental results were not provided to demonstrate the findings[17]. A broadband rectenna consisting of a dual-circularly polarized spiral rectenna array operating over a frequency range of 2–18 GHz was demonstrated[18]. The rectified DC power was characterized as a function of DC load, RF frequency and polarization for power densities between 10^{-5} and 10^{-1} mW/cm^2. However, the proposed rectenna was matched at a single input RF power level for a specified load resistance for the characterization. Also, due to the low Q value of the rectifier circuit, the conversion efficiency was a fraction of 1% at -15.5 dBm. From a design point of view, while it is relatively easy to achieve a broadband antenna, it is very challenging to realize a broadband rectenna due to the non-linearity of the rectifier impedance with input power across the frequency band[19].

To address this, a promising approach is to use a dual-band or multi-band configuration. This can maximize the power conversion efficiency (PCE) at the specific frequencies where the maximum ambient signal level is

available. Various dual-band RF energy harvesting systems has been demonstrated[20–24], however a large signal analysis of the rectifier was commonly not provided over a broad input power range. A dual-band RF energy harvesting using frequency limited dual-band impedance matching has been proposed[20] and the PCE was shown over a high power range of 0 to 160 mW, however it was only matched at a single input power level (10 dBm). A CMOS dual-narrowband energy harvester circuit was modeled at environmental power levels[21]. Again, the rectifier efficiency was demonstrated with only single input power levels of −19 and −19.3 dBm at 2 GHz and 900 MHz respectively, and a large signal analysis was not presented. A compact dual-band rectenna operating at 915 MHz and 2.45 GHz has been demonstrated[22] and the PCE was shown for input power levels of −15, −9 and −3 dBm. However, the reflection coefficient was evaluated at a single incident power level. Furthermore, the efficiency results with dual-tone excitation simultaneously and single-tone excitation (at 915 MHz) are very similar, hence the impact of applying a dual-band technique does not demonstrate a clear advantage over a single band. A dual-frequency rectenna for WPT has been proposed[23] which achieved a conversion efficiency of 84.4% and 82.7% at 2.45 and 5.8 GHz with a high input power level of 89.84 and 49.09 mW respectively. These power levels far exceed ambient levels in the environment[19]. A conformal hybrid solar and electromagnetic (EM) energy harvesting rectenna has been presented[24] and the PCE was provided with −30 to 5 dBm input power, achieving an efficiency up to 40% at 1.85 GHz for higher input power levels (above −5 dBm). However the reflection coefficient was not provided at low input power range.

A multi-resonant rectenna that uses a multi-layer antenna and rectifier has been evaluated for a −16 dBm to +8 dBm RF received power level, but the rectifier circuit layout and large signal analysis were not provided to clarify the findings[25]. Furthermore, a rectenna for triple-band biotelemetry communications has been proposed using a triple-band antenna and single frequency rectifier[26]. However, this rectenna is not suitable for RF energy scavenging due to the low efficiency at lower input power levels. Another triple band rectenna presented an RF-DC efficiency over the input power range of −14 to +20 dBm[27], however the reflection coefficient results were only evaluated at a single input power level. This rectenna was shown to harvest 7.06 µW of DC power from three sources simultaneously at a high input power level of +10 dBm. A multi-band harvesting system has also been proposed where four individual harvesters are designed to cover four frequency bands[28]. However, a large signal analysis was not provided over a broad input power range. Furthermore, the proposed harvesting system has a minimum sensitivity of −25 dBm, whilst in a real environment more sensitive systems are required as the available RF power levels are very low[19].

Tunable impedance matching networks have been demonstrated in order to collect RF signals from various sources and convert them to DC power[29]. However from an application point of view, this is still single frequency rectification and it is not widely applicable to environmental RF energy scavenging where the available power is very low.

In order to increase the amount of RF energy scavenged by a rectenna, it is crucial to identify and harvest multiple ambient frequency sources over their realistic available energy range. Our previous research has demonstrated the feasibility of RF energy harvesting through RF field investigations and maximum available power analysis in metropolitan areas of Melbourne, Australia[19]. The maximum available power for different frequency bands based on antenna aperture and number of antennas in a given collection area was analyzed. Measured results and analysis indicated that cellular systems and broadcast sources are well suited to harvesting, with scavengeable RF power ranging from −40 to −10 dBm. This identifies two important considerations in the design of efficient rectenna

for RF energy harvesting: the scavengeable ambient RF power sources available, and the significant variance of this power.

The RF to DC rectifier solutions proposed in recent literature have focused on maximizing the system efficiency at a given, and often quite high, input power level. This neglects the issues related to input power variation which can lead to unexpected variations in the matching network due to diode non-linearity. Also, the scavangeable levels of ambient RF power have been shown to be orders of magnitude lower. Therefore, based on our previous research outcomes and recommendations[19], an efficient power harvesting solution could encompass a multi-band matching circuit at the specific frequencies where maximum signal power is available, enabling greater power harvesting due to the combination of RF signals. This also results in a higher power being fed to a single rectifier, utilizing the diode function more efficiently.

This paper presents an RF energy harvesting method that can scavenge a wide range of ambient power levels which are orders of magnitude lower than previous reported techniques in the literature. An efficient dual resonant rectifier circuit is proposed, matched to a 50 Ω input port at 490 and 860 MHz over a broad low input RF power range from −40 to −10 dBm. The proposed dual resonant matching network operates efficiently at two identified harvesting frequency bands over a wide input power range, maximizing DC power by scavenging two sources simultaneously.

The remainder of this paper is organized as follows. First, the key results for the reflection coefficient and output DC power are presented. Subsequently, the Discussion section summarizes the results and demonstrates their potential implications, the limitations of this study, open questions and future research. Finally, the Method section describes the proposed rectifier design.

Results

A dual resonant rectifier was fabricated on a 1.58 mm FR-4 substrate with a dielectric constant $\varepsilon_r \approx 4.5$ and a loss tangent $\delta \approx 0.025$. These substrate parameters were measured using the Nicolson-Ross method[30] so accurate values could be used in the rectifier design. A photograph of the fabricated dual resonant rectifier is shown in Fig. 1 which depicts input RF port, dual-band matching network lumped components, Schottky diodes and the output terminal. The performance of the rectifier was verified by measuring the input reflection properties, and the output power was calculated from the measured output DC voltage for the input powers from −40 to −10 dBm.

Reflection Coefficient. The $|S_{11}|$ of the rectifier was evaluated using a vector network analyzer (VNA). The VNA was re-calibrated for each input power level. Figure 2 compares the simulated and measured $|S_{11}|$ versus frequency for the dual resonant rectifier circuit at four different input power levels from −40 to −10 dBm. The measured results show very good agreement to the simulations. Slightly higher reflection was observed for the resonant frequencies at the lower part of the input power range (due to the diode characteristics). However, the proposed rectifier circuit is well-matched ($|S_{11}| < −10$ dB) at the desired frequency bands of
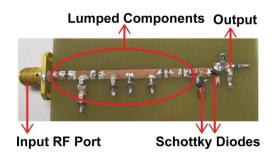
Figure 1 | Fabricated rectifier prototype.
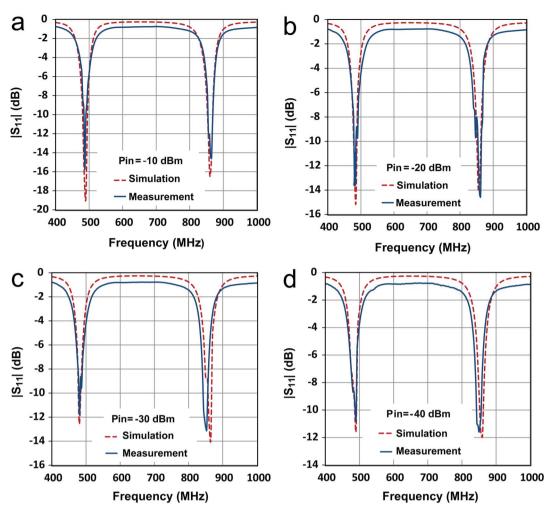
Figure 2 | Simulated and measured $|S_{11}|$ as a function of frequency and input RF power for the proposed dual resonant rectifier circuit. (a) −10 dBm. (b) −20 dBm. (c) −30 dBm. (d) −40 dBm.

478–496 MHz and 852–869 MHz over the broad range of input powers from −40 to −10 dBm. The small difference between simulation and measurement is due to the parasitic extraction accuracy.
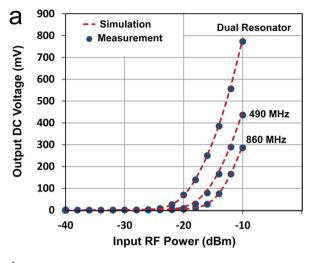
Output DC Power. In the frequency domain, the Harmonic Balance method of analysis provides a comprehensive treatment of a multispectral problem[18]. The method intrinsically takes into account the DC component and a specified number of harmonics, while allowing the ability to specify the source impedance and harmonic terminations. A Harmonic Balance simulation was used to numerically evaluate the output DC voltage of the dual resonant rectifier for both a single and two tone input. The output DC voltage across the load resistor was also measured and used to calculate output DC power. Measurements were performed using a Wiltron 68247B synthesized signal generator as a RF power source for the rectifier circuit. Recording of the output DC voltage across the load resistance was achieved with a Fluke 79III digital voltage meter. The RF source power was initially set at −10 dBm, and decreased in 2 dB steps. In the dual-band measurement case, two RF signal generators were fed to the rectifier circuit simultaneously via a power combiner.

The simulation and measurement results for single and dual input tones are summarized in Fig. 3(a) and (b). A measured DC voltage of 772.8 mV is achieved with two simultaneous input tones at an input power of −10 dBm. For single tone measurements, DC voltages of 436 mV and 286 mV at 490 MHz and 860 MHz respectively are produced. The comparison between the 490 and 860 MHz single rectifiers highlights the impact of the input frequency on the PCE.

A higher amount of DC voltage can be generated at the lower frequency. This difference comes from decreasing diode performance at the higher frequency due to the higher junction capacitance of the diode[31].

Importantly, a slightly higher DC voltage can be generated with the dual resonant rectifier as compared to the sum of output voltage from the two single bands, particularly at the lower input power levels as can be seen in Fig. 3(b) which shows the lower power section of Fig. 3(a) in more detail. By maximizing power collection from various sources of different frequencies and delivering the combined power to the rectification circuit, the diode conversion efficiency is enhanced which results in a higher level of rectified voltage. Figure 4 compares the simulated and measured output DC power for the dual resonant rectifier circuit with both single and dual input tones. A measured DC power of 17.3 μW and 7.5 μW can be generated at 490 MHz and 860 MHz respectively with a single tone input of −10 dBm (100 μW). This represents true efficiencies of 17.3% and 7.5% for the individual single band rectification (see Fig. 5). However, the measured DC output power with two concurrent input tones of −10 dBm is 54.3 μW which corresponds to an effective efficiency of 54.3% for the dual-band rectifier (see Fig. 6). This represents a 3.14 and 7.24 times increase in output DC power over the single tone excitation at 490 MHz or 860 MHz respectively. This trend is evident down to low input power levels (around 40 μW). Furthermore, there is a significant increase in the PCE of the dual resonant rectifier for lower input power levels (<40 μW).
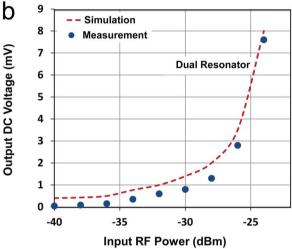
Figure 3 | Output DC voltage as a function of input RF power for single input tone at both 490 MHz and 860 MHz and for dual input tones (a) with −40 to −10 dBm input RF power (b) with −40 to −25 dBm input RF power. (This power range is associated with the signal source).
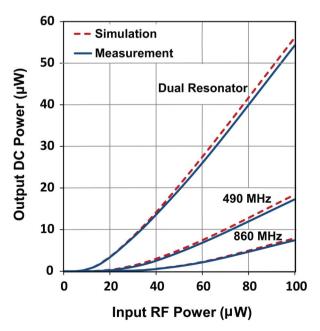
Figure 4 | Output DC power as a function of input RF power for single and dual input tones.
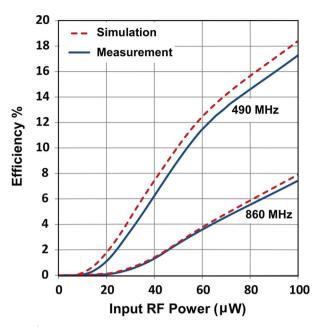
Figure 5 | RF to DC conversion efficiency as a function of input RF power for single band rectification.

Here, the effective efficiency is defined as the ratio of output DC power to the available input RF power rather than the power delivered to the diodes (equation (1)). The available power level is associated with the signal source. For the single resonator, the available input power is −10 dBm and the delivered power is also −10 dBm (assuming no loss). However, by creating a dual resonant matching network the power delivered to the diodes is −7 dBm (combined total input power from two signal generators) but the available power is still −10 dBm.

$$\eta = \frac{P_o}{P_i} = \frac{\left(\frac{V_o{}^2}{R_L} \right)}{P_i} \tag{1}$$

Therefore, combining input RF signals into a single rectification stage results in high sensitivity rectifier, which is widely applicable to
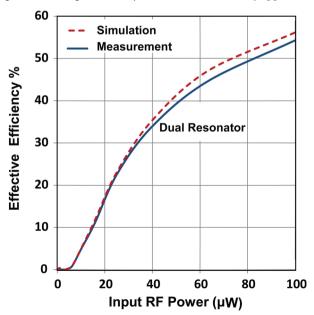
Figure 6 | Effective RF to DC conversion efficiency as a function of input RF power for dual resonant rectification.

Table 1 | Rectifier performance comparison

Ref.	Technology	Measured Efficiency (%)	RF power variation (in PCE evaluation)	Rectification Technique
13	Schottky diode	82@50 mW	N/A	Single resonator
16	Schottky diode	44@−10 dBm	N/A	Single resonator
18	Schottky diode	20@0.07 mW/cm² 0.1@5×10⁻⁵ mW/cm²	10⁻⁵ to 10⁻¹ mW/cm²	Broad band
20	Schottky diode	77.13@22 dBm (158.49 mW)	0 to 160 mW	Dual resonator
21	CMOS	9.1@−19.3 dBm (900 MHz) 8.9@−19 dBm (2 GHz)	N/A	Dual resonator
22	Schottky diode	37 (915 MHz)@−9 dBm 30 (2.45 MHz)@−9 dBm	−40 to 0 dBm	Dual resonator
23	Schottky diode	84.4@89.84 mW (2.45 GHz) 82.7@49.09 mW (5.8 GHz)	0 to 100 mW	Dual resonator
27	Schottky diode	80@10 dBm (940 MHz) 47@8 dBm (1.95 GHz) 43@16 dBm (2.44 GHz)	−14 to 20 dBm	Triple resonator
29	Schottky diode	50@−5 dBm	−25 to 0 dBm	Dual resonator with tunable input response
This work	Schottky diode	54.3@−10 dBm 11.25@−18 dBm (490 and 860 MHz)	−40 to −10 dBm	Dual resonator

Table 2 | Environmental measurement results

Suburb	Available frequencies (MHz)	Respective available RF power (dBm) [μW]	Measured DC power (μW)
Bayswater	486, 488, 489, 490, 491, 867, 868, 869, 870, 871, 872, 873, 874	−19[12.5], −20[10], −17[19.95], −15[31.62], −22[6.3], −37[0.199], −37[0.199], −30[1], −24[3.98], −20[10], −30[1], −37[0.199], −40[0.1]	39.38
Bentleigh	491, 492, 494, 495, 865, 866, 867, 868, 869, 870, 871	−12[63.09], −46[0.02], −42[0.063], −57[0.001], −27[1.99], −27[1.99], −30[1], −37[0.199], −40[0.1], −40[0.1], −41[0.07]	30.9
RMIT University (Melbourne CBD)	487, 488, 489, 490, 491, 851, 861, 862, 866, 867, 868, 869	−30[1], −22[6.3], −29[1.25], −22[6.3], −20[10], −23[5.01], −21[7.94], −21[7.94], −30[1], −35[0.31], −40[0.1], −40[0.1]	14.5

real environmental RF energy scavenging. This multi-band technique can provide higher DC power than combining two separate single frequency rectifier circuits operating at the same frequencies. This is due to the fact that harvesting RF energy from various available sources simultaneously increases the delivered power to the rectifier, which improves the diode conversion efficiency and consequently enhances the output DC power. Table 1 summarizes this work as compared to previous published work.

In order to provide a realistic scenario for the proposed dual-band rectifier, measurement results were taken in three suburbs of Melbourne, Australia, congruent with our previous research outcomes[19]. Table 2 summarizes these environmental measurement results. It should be noted that the lower band (478–496 MHz) has a 3.67% fractional bandwidth and the higher band (852–869 MHz) has around 2% fractional bandwidth. Hence, various RF frequencies from different sources can be harvested within these two bands. The environmental measurement results demonstrate the feasibility of harvesting ambient EM energy from multiple sources simultaneously.

Discussion

The feasibility of harvesting ambient EM energy from multiple sources simultaneously is investigated in this paper. The proposed dual resonant rectifier operates at two frequency bands (478–496 and 852–869 MHz), which are used for broadcasting and cellular systems respectively. The dual resonant rectifier exhibits favorable impedance matching over a broad input power range (−40 to −10 dBm) at these two bands. The achieved sensitivity and dynamic range demonstrate the usefulness of this innovative low input power rectification technique. Simulation and experimental results of input reflection coefficient and rectified output power are in excellent agreement. The measurement results demonstrate that a two tone input to the proposed dual-band RF energy harvesting system can generate 3.14 and 7.24 times more power than a single tone at 490 or 860 MHz respectively, resulting in a measured effective efficiency of 54.3% for a dual-tone input power of −10 dBm. It is evident that this dual resonant rectification technique increases the RF to DC effective conversion efficiency, and hence the recoverable DC power for low power applications. Furthermore from a design and economic perspective, utilizing a large number of components (e.g. antennas, diodes) to realize individual rectifier circuits for each frequency band creates additional expense. In order to provide more realistic measurement results, the proposed dual-band rectifier was tested in three suburbs of Melbourne, Australia. Therefore, this dual-band technique offers a simple and cost-effective solution which is of paramount importance for environmental power harvesting systems. This innovative technique has the potential to generate a viable perpetual energy source for low power applications in urban environments.

Limitation of the study, open questions and future work

Utilizing diodes which are more suitable to low power applications ($P_i <$ −20 dBm) could increase the voltage sensitivity, resulting in a higher RF-DC conversion efficiency[32]. Applying a power optimized waveform excitation to the rectifier circuit in these frequency bands, a higher amount of DC power can be generated when compared with a single and dual tone excitations with the same input power[33,34]. However this technique is not applicable to energy harvesting where the input waveform is arbitrary.

Utilizing our proposed dual resonant rectification technique to combine resonant circuits for any other arbitrary frequency bands could lead to PCE improvement, provided that suitable diodes for the desired frequency bands are selected. Note that by increasing the operating frequency the rectification performance degrades due to
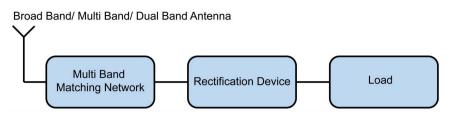
Figure 7 | General block diagram of the RF energy harvesting system.

the higher junction capacitance of the diode. Hence, lower output voltage is expected at higher frequency bands.

It is the object of our future work to design a multi-band rectenna array for enhanced RF energy harvesting. Furthermore, increasing the bandwidth, sensitivity and efficiency will also be investigated.

Methods

The major goal in designing an efficient RF harvesting system is to produce high DC output power. Toward this goal, a high sensitivity rectifier is crucial for optimum RF scavenging. A significant factor governing the sensitivity of a rectifier is the threshold voltage of the diode used for rectification. The diode must be able to "switch on" for very low ambient energy levels.

To address this sensitivity issue, a system that scavenges power from multiple frequency bands and combines them to activate a rectification circuit is proposed. The general block diagram of the proposed system is depicted in Fig. 7. Various environmental RF energy sources of different frequencies are collected by an appropriately designed antenna, and delivered to the rectification circuit via a multi-band matching network. The rectification circuit converts the combination of RF signals into DC power for low-power applications. The embodiment in this paper realizes a dual resonant matching circuit as a transition between a 50 Ω nominal antenna output and the non-linear rectification device at 490 and 860 MHz. Based on the Australian Radiofrequency Spectrum Plan[35], these bands are allocated to broadcasting services and cellular systems.

Device Selection. Due to the very low ambient power available in a real environment[19], a very low threshold voltage rectification device is required in order to increase sensitivity. For this reason, Schottky diodes (GaAs or Si) are commonly employed for RF energy harvesting. In this work, a microwave Schottky detector HSMS2820 ($C_{j0} = 0.7$ pF, $R_s = 6$ Ω, $I_s = 2.2e^{-8}$ A) is chosen due to its excellent high frequency performance, low series resistance (R_s) and junction capacitance (C_j), and low threshold voltage with high-saturation current[31]. This low threshold voltage (0.15–0.3 V) supports rectification at low input power levels.

Proposed Rectifier Design. In order to design an efficient RF harvesting system, the non-linearity of the rectifier impedance with frequency and input power should be matched to the 50 Ω output of the antenna at the desired frequency bands. Therefore, the diode input impedance as a function of frequency and different power levels were calculated and analyzed[36]. In order to match the input impedance of the rectifier to the 50 Ω output of the antenna, the total load impedance for different input power and frequencies should be determined. A circuit consisting of a pair of Schottky Barrier Diodes (SBD) terminated with a load resistor ($R_{Load} = 11$ kΩ) and an output bypass capacitor ($C2 = 6.8$ pF) was simulated using Agilent ADS software. Figure 8 shows the proposed geometry of the voltage-doubler topology[31,37]. The voltage doubler rectifier structure is employed for the design of the RF-DC power conversion system as this topology is well suited to low power rectification. The resistor and capacitor at the output will filter high frequencies. The high load resistor (11 kΩ) was chosen to observe a reasonable output voltage at very low currents. Using Large Signal S-Parameters analysis in Agilent ADS software, the load impedance and bypass capacitor were determined and optimized.

The voltage doubler rectifier in Fig. 8 consists of a peak rectifier formed by D2 and bypass capacitor C2 (6.8 pF) and a voltage clamp formed by D1 and C1 (total capacitance of the transmission lines and diode's parasitic capacitance (C_p)). In the
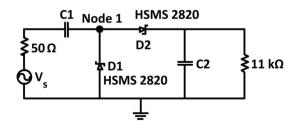
negative phase of the input, current flows through D1 while D2 is cutoff. The voltage across D1 stays constant around its threshold voltage and the voltage at node 1 is charged to $-V_{th1}$ (where $-V_{th1}$ is the threshold voltage of the D1). At the negative peak, the voltage across C1 is $V_{amp} - V_{th1}$ (where V_{amp} is the amplitude of the input signal). In the positive phase of the input, current flows through D2 while D1 is cutoff. The voltage across C1 remains the same as the previous phase because it has no way to discharge. At the positive peak, the voltage across D1 is $2V_{amp} - V_{th1}$. Since D2 is conducting current to charge C2, the voltage at the output is $V_{out} = 2V_{amp} - V_{th1} - V_{th2}$.

The DC equivalent circuit of the SBD is a voltage source in series with the junction resistor R_j which is obtained by differentiating the diode voltage–current characteristic and is given by equation (2)[31,38]:

$$R_j = \frac{nKT}{q(I_s + I_b)} \qquad (2)$$

Where n is the diode ideality factor, K is the Boltzmann's constant, T is the temperature in degrees Kelvin, q is the electronic charge, I_s is the diode saturation current and I_b is the external bias current. At low power levels, the saturation current is very small ($I_s = 2.2$ e^{-8} A) and for a zero-biased diode, $I_b = 0$. Therefore, the resulting value of junction resistance at room temperature is approximately 1.7 MΩ. Since, the saturation current is highly temperature dependent, R_j will be even higher at lower temperatures which tends to decrease the output voltage. As the input power increases, some circulating rectified current will cause a drop in the value of R_j and this phenomenon will increase the value of the DC output voltage. Furthermore, it is worth to highlight that the rectified current produced by the first diode (D1) in Fig. 8 constitutes the external bias current of the second diode (D2) which will help to reduce the R_j and hence the detection sensitivity is improved. Therefore, depending on the amount of available bias current, R_j is varying (equation (2)), hence the matching network is changing which impacts the amount of delivered power to the diode and results in different values of PCE.

A Schottky barrier diode can be modeled by the linear equivalent circuit shown in Fig. 9, where L_p and C_p are the diode's parasitic inductance and capacitance respectively due to packaging ($L_p = 2$ nH and $C_p = 0.08$ pF) which are generally unwanted[39]. This linear model is used for determining the diode impedance at a given input power.

The diode impedance analyzed using a Harmonic Balance simulator and a non-linear model of the diodes over the frequency range of 400 to 900 MHz at various input power levels (Fig. 10). Due to the large junction resistor at low input RF power levels, the rectification device is turned off in absence of an appropriate matching network. Large Signal S-parameter analysis was conducted and higher input power (associated with the signal source) is applied directly to the Schottky diodes configuration of Fig. 8 which does not include a matching network in order to turn on the diodes (reduce the value of R_j) and extract approximate input impedance value as our starting point in design of a matching network. As it can be seen in Fig. 10, with increasing the source power, the diode impedance is varying and it is beginning to switch on. Hence, the input impedance needs to be determined when the diode is

Figure 8 | Schematic of a voltage-double rectifier without matching network.
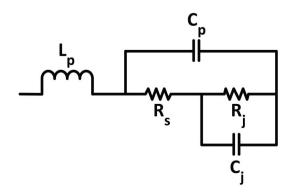
Figure 9 | HSMS 2820 Schottky diode equivalent circuit.
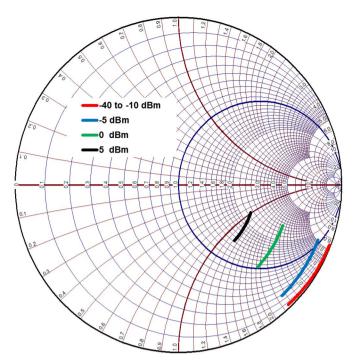
turned on to realize the matching network for a rectifier circuit. Obviously, in the presence of an appropriate matching network the rectification device can be turned on at lower power levels, whilst in the absence of a matching network a higher input power should be applied to switch on the diode. (Note that, with an unmatched rectifier the total applied input power from the signal source cannot be delivered to the diode due to the high reflection in the circuit).

The aim is to match the input impedance of the device to 50 Ω at 478–496 MHz and 852–869 MHz bands over a broad range of input RF powers. The procedure commences by matching the diode input impedance at high unmatched source power and shifting the diode input impedance at various power levels to within the voltage standing wave ratio (VSWR) <2 circle on the Smith chart. This procedure assumes that diode input impedance does not drastically change in this low power range. The simulation results of Fig. 10 prove that this is the case.

In order to provide maximum power transfer from the antenna to the rectifier circuit, a dual resonant rectifier network is designed as a transition between a 50 Ω nominal antenna output and the non-linear rectification device over the power range of −40 to −10 dBm (see Fig. 11). Hence, a coupled-resonator structure with both series and shunt resonators is designed to achieve a dual-band network[40]. The linear equivalent circuit model of the SBD chip[39] has been taken into consideration to design the dual band match at the desired frequency bands. In Fig. 11, $C_{equivalent}$ represents the total capacitance of the diodes and bypass capacitor and $L_{equivalent}$ is the overall parasitic inductance of the diodes. The series L-C resonator ($L4 + L_{equivalent}$ and $C_{equivalent}$) and the parallel L-C resonator ($C3$ and $L3$) define the dual resonant circuit. The series resonator corresponds closely to the higher band specification of 852–869 MHz, whilst the parallel resonator approximates the lower 478–496 MHz band. A minimum number of components were used in order to reduce the ohmic and parasitic losses.

The resonant frequency of each sub-circuit was determined in isolation using the following equation:

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{3}$$

The 852–869 MHz band resonator circuit components were calculated. Here, $C = C_{equivalent} \cong 1.3$ pF consists of the combination of the bypass capacitor (6.8 pF) and

Figure 10 | Diode input impedance calculated with Large Signal S-parameter analysis over the frequency range of 400 to 900 MHz with various unmatched input power levels.

Figure 11 | Schematic of a dual resonant rectifier (optimized parameters of the chip components are: $L1 = 3.9$ nH, $C1 = 0.2$ pF, $L2 = 12$ nH, $C2 = 1.8$ pF, $L3' = 3.9$ nH, $C3' = 7.5$ pF, $L4' = 11.6$ nH, $L_{equivalent} \cong 1$ nH, $C_{equivalent} \cong 1.3$ pF).
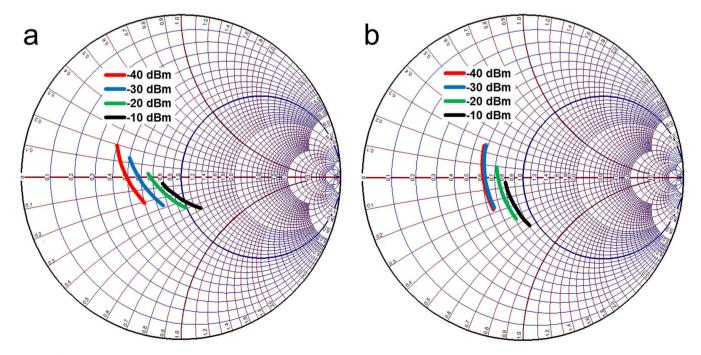
Figure 12 | Dual resonant impedance matching with −40 to −10 dBm input RF power. (a) 478–496 MHz. (b) 852–869 MHz.

the overall junction ($C_{j0} = 0.7$ pF) and parasitic capacitance ($C_p = 0.08$ pF) of $D1$ and $D2$. Thus, L is calculated to be 26.5 nH in order to achieve an appropriate resonant frequency. Note that, L consists of L_4 and the overall parasitic inductance ($L_{equivalent} \cong$ 1 nH) of $D1$ and $D2$. The 478–496 MHz band resonator circuit components were calculated as $C3 \cong 15$ pF and $L3 \cong 7.2$ nH. Hence, the initial component values are determined for the two resonant circuits.

Initially these resonators were combined to achieve a dual-band structure. Then standard LC matching technique[40] is utilized to determine $C1$, $C2$, $L1$, and $L2$ to achieve minimum reflection at the resonant frequencies. The substitution of realistic chip component values with their associated parasitics, and addition of 50 Ω microstrip lines and T-junctions introduce delay and shift the imaginary part of the input impedance. The via-holes also contribute to extra inductance in the circuit. Hence minor circuit adjustments are made in order to fine tune the resonant frequencies to the desired values. The final optimized values of the standard chip components are: $L3' = 3.9$ nH, $C3' = 7.5$ pF and $L4' = 11.6$ nH. Large Signal S-parameter analysis is also performed to demonstrate the matching network performance as the input power is varied. Simulation results for the input impedance of the circuit depicted in Fig. 11 are illustrated in Fig. 12. The proposed dual-resonant matching circuit achieves a VSWR <2 at 478–496 MHz and 852–869 MHz for input power ranging from −40 to −10 dBm. It should be noted that the matching circuit was designed based on the input impedance of two diodes and the output resistor and capacitor (Fig. 10). Therefore, selecting a different value for the load resistor requires a new matching circuit to be designed.

1. Paradiso, J. A. & Starner, T. Energy scavenging for mobile and wireless electronics. *IEEE Pervasive Comput.* **4**, 18–27 (2005).
2. Singh, R., Gupta, N. & Poole, K. F. Global green energy conversion revolution in 21st century through solid state devices. 26th International Conference on Microelectronics, Nis, Serbia. *IEEE*. DOI:10.1109/ICMEL.2008.4559221. 45–54 (11–14 May, 2008).
3. Wang, Y. K., Sheng, G. P., Shi, B. J., Li, W. W. & Yu, H. Q. A Novel electrochemical membrane bioreactor as a potential net energy producer for sustainable wastewater treatment. *Sci. Rep.* **3**, DOI:10.1038/srep01864 (2013).
4. Inayat, S. B., Rader, K. R. & Hussain, M. M. Nano-materials enabled thermoelectricity from window glasses. *Sci. Rep.* **2**, DOI:10.1038/srep00841 (2012).
5. Brown, W. C. The history of power transmission by radiowaves. *IEEE Trans. Microw. Theory Tech.* **32**, 1230–1242 (1984).
6. Le, T., Mayaram, K. & Fiez, T. Efficient far-field radio frequency energy harvesting for passively powered sensor networks. *IEEE J. Solid-State Circuits.* **43**, 1287–1302 (2008).
7. Paolo, J. & Gaspar, P. D. Review and future trend of energy harvesting methods for portable medical devices. Proceedings of the World Congress on Engineering (WCE). London, U. K. 909–914 (30 June–2 July, 2010).
8. Matsuki, H., Yamakata, Y., Chubachi, N., Nitta, S. & Hashimoto, H. Transcutaneous DC-DC converter for totally implantable artificial heart using synchronous rectifier. *IEEE Trans. Magn.* **32**, 5118–5120 (1996).
9. Scorcioni, S. *et al.* RF to DC CMOS rectifier with high efficiency over a wide input power range for RFID applications. Microwave Symposium Digest (IMS), *IEEE MTT-S.* Montreal, QC, Cabada. DOI:10.1109/MWSYM.2012.6259760. 1–3 (17–22 June, 2012).
10. Shameli, A., Safarian, A., Rofougaran, A., Rofougaran, M. & De Flaviis, F. Power harvester design for passive UHF RFID tag using a voltage boosting technique. *IEEE Trans. Microw. Theory Tech.* **55**, 1089–1097 (2007).
11. Mandal, S. & Sarpeshkar, R. Low-power CMOS rectifier design for RFID applications. *IEEE Trans. Circuits Syst. I, Reg. Papers.* **54**, 1177–1188 (2007).
12. Sun, H., Guo, Y. X., He, M. & Zhong, Z. Design of a high-efficiency 2.45-GHz rectenna for low-input-power energy harvesting. *IEEE Antennas Wireless Propag. Lett.* **11**, 929–932 (2012).
13. McSpadden, J. O., Fan, L. & Chang, K. Design and experiments of a high-conversion-efficiency 5.8-GHz rectenna. *IEEE Trans. Microw. Theory Tech.* **46**, 2053–2060 (1998).
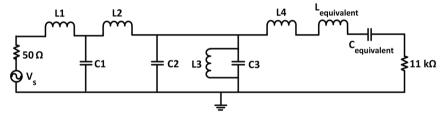14. Din, N. M., Chakrabarty, C. K., Bin Ismail, A., Devi, K. K. A. & Chen, W. Y. Design of RF energy harvesting system for energizing low power devices. *PIER.* **132**, 49–69 (2012).
15. Barnett, R. E., Liu, J. & Lazar, S. A RF to DC voltage conversion model for multi-stage rectifiers in UHF RFID transponders. *IEEE J. Solid-State Circuits.* **44**, 354–370 (2009).
16. Kanaya, H. *et al.* Energy harvesting circuit on a one-sided directional flexible antenna. *IEEE Microw. Wireless Compon. Lett.* **23**, 164–166 (2013).
17. Buonanno, A., D'Urso, M. & Pavone, D. An ultra wide-band system for RF energy harvesting. 5th European Conference on Antennas and Propagation (EUCAP). Rome, Italy. *IEEE.* 388–389 (11–15 April, 2011).
18. Hagerty, J. A., Helmbrecht, F. B., McCalpin, W. H., Zane, R. & Popovic, Z. B. Recycling ambient microwave energy with broad-band rectenna arrays. *IEEE Trans. Microw. Theory Tech.* **52**, 1014–1024 (2004).
19. Shariati, N., Rowe, W. S. T. & Ghorbani, K. RF field investigation and maximum available power analysis for enhanced RF energy scavenging. 42nd European Microwave Conference (EuMC). Amsterdam RAI, The Netherlands. *IEEE.* 329–332 (29 October–1 November, 2012).
20. Kim, P., Chaudhary, G. & Jeong, Y. A dual-band energy harvesting using frequency limited dual-band impedance matching. *PIER.* **141**, 443–461 (2013).
21. Li, B. *et al.* An antenna co-design dual band RF energy harvester. *IEEE Trans. Circuits Syst. I, Reg. Papers.* **60**, 3256–3266 (2013).
22. Niotaki, K. *et al.* A compact dual-band rectenna using slot-loaded dual band folded dipole antenna. *IEEE Antennas Wireless Propag. Lett.* **12**, 1634–1637 (2013).
23. Suh, Y. H. & Chang, K. A high-efficiency dual-frequency rectenna for 2.45- and 5.8-GHz wireless power transmission. *IEEE Trans. Microw. Theory Tech.* **50**, 1784–1789 (2002).
24. Collado, A. & Georgiadis, A. Conformal hybrid solar and electromagnetic (EM) energy harvesting rectenna. *IEEE Trans. Circuits Syst. I, Reg. Papers.* **60**, 2225–2234 (20013).
25. Rizzoli, V., Bichicchi, G., Costanzo, A., Donzelli, F. & Masotti, D. CAD of multi-resonator rectenna for micro-power generation. European Microwave Integrated Circuits Conference (EuMIC). Rome, Italy. *IEEE.* 331–334 (228–29 September, 2009).
26. Huang, F. J. *et al.* Rectenna application of miniaturized implantable antenna design for triple-band biotelemetry communication. *IEEE Trans. Antennas Propag.* **59**, 2646–2653 (2011).
27. Pham, B. L. & Pham, A. V. Triple bands antenna and high efficiency rectifier design for RF energy harvesting at 900, 1900 and 2400 MHz. Microwave Symposium Digest (IMS), *IEEE MTT-S.* Seattle, WA. DOI:10.1109/MWSYM.2013.6697364. 1–3 (2013).
28. Pinuela, M., Mitcheson, P. D. & Lucyszyn, S. Ambient RF energy harvesting in urban and semi-urban environments. *IEEE Trans. Microw. Theory Tech.* **61**, 2715–2726 (2013).
29. Mikeka, C. & Arai, H. Microwave Tooth for sensor power supply in battery-free applications. Proceedings of the Asia-Pacific Microwave Conference (APMC). Melbourne, VIC. *IEEE.* 1802–1805 (5–8 December, 2011).
30. Baum, T., Thompson, L. & Ghorbani, K. Complex dielectric measurements of forest fire ash at X-band frequencies. *IEEE Geosci. Remote Sens. lett.* **8**, 859–863 (2011).
31. Avago Technologies, *Surface Mount RF Schottky Barrier Diodes HSMS-282X Series. Data sheet.* (2005). Available at: http://www.avagotech.com/docs/AV02-1320EN. (Accessed: 22/11/2014).
32. Avago Technologies, *Surface Mount Zero Bias Schottky Detector Diodes HSMS-285X Series. Data sheet.* (2005). Available at: http://www.avagotech.com/docs/AV02-1377EN. (Accessed: 22/11/2014).
33. Valenta, C. R. & Durgin, G. D. Rectenna performance under power-optimized waveform excitation. IEEE International Conference on RFID. Penang. DOI:10.1109/RFID.2013.6548160. 237–244 (30 April–2 May, 2013).
34. Trotter, M. S. & Durgin, G. D. Survey of range improvement of commercial RFID tags with power optimized waveforms. IEEE International Conference on RFID. Orlando. DOI:10.1109/RFID.2010.5467265. 195–202 (14–16 April, 2010).
35. Australian Government, *Australia Radio Frequency Spectrum Plan (ACMA).* (January 2013). Available at: http://acma.gov.au/~/media/Spectrum%20Transformation%20and%20Government/Information/pdf/Australian%20Radiofrequency%20Spectrum%20Plan%202013.pdf. (Accessed: 22/09/2014).
36. Keyrouz, S., Visser, H. J. & Tijhuis, A. G. Rectifier analysis for radio frequency energy harvesting and power transport. 42nd European Microwave Conference (EuMC). Amsterdam RAI, The Netherlands. *IEEE.* 428–431 (29 October–1 November, 2012).
37. Avago Technologies, *Schottky Diode Voltage Doubler. Application Note 956-4.* (2005). Available at: http://c1233384.r84.cf3.rackcdn.com/UK_HPA_HSCH-5331_7AN.pdf. (Accessed: 22/11/2014).
38. Eriksson, H. & Waugh, R. W. *A temperature compensated linear diode detector. Design Tip.* Agilent Technologies. (2000). Available at: http://electronix.ru/forum/index.php?act=Attach&type=post&id=13726. (Accessed: 22/11/2014).
39. Avago Technologies, *The Zero Bias Schottky Diode Detector at Temperature Extremes-Problems and Solutions. Application Note AN 1090.* (2005). Available at: http://www.efo.ru/download/Zero_bias_shottky_over_temperature.pdf. (Accessed: 22/11/2014).
40. Mattaei, G., Young, L. & Jones, E. M. T. *Microwave Filters, Impedance-Matching Networks and Coupling Structures.* Artech house books, Dedham, MA. (1980).

Acknowledgments

Author contributions
N.S. designed and simulated the rectifier, carried out the measurements, interpreted results and wrote the paper. K.G. directed the research, contributed to perform the simulation and measurement and validation of design and results. W.S.T.R. supervised the research, analyzed the data and contributed to the general concept, validation of design and results. J.R.S. analyzed the data and contributed to the general concept, validation of design and results. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

sensors

Article

Optimization of Passive Low Power Wireless Electromagnetic Energy Harvesters

Antwi Nimo *, Dario Grgić and Leonhard M. Reindl

Department of Microsystems Engineering, Laboratory for Electrical Instrumentation,
University of Freiburg, IMTEK, Georges-Köhler-Allee 106, 79110 Freiburg, Germany;
E-Mails: grgic@imtek.de (D.G.); leonhard.reindl@imtek.uni-freiburg.de (L.M.R.)

* Author to whom correspondence should be addressed; E-Mail: antwi.nimo@imtek.de;
Tel.: +49-761-203-7223; Fax: +49-761-203-7222.

Abstract: This work presents the optimization of antenna captured low power radio frequency (RF) to direct current (DC) power converters using Schottky diodes for powering remote wireless sensors. Linearized models using scattering parameters show that an antenna and a matched diode rectifier can be described as a form of coupled resonator with different individual resonator properties. The analytical models show that the maximum voltage gain of the coupled resonators is mainly related to the antenna, diode and load (*remote sensor*) resistances at matched conditions or resonance. The analytical models were verified with experimental results. Different passive wireless RF power harvesters offering high selectivity, broadband response and high voltage sensitivity are presented. Measured results show that with an optimal resistance of antenna and diode, it is possible to achieve high RF to DC voltage sensitivity of 0.5 V and efficiency of 20% at −30 dBm antenna input power. Additionally, a wireless harvester (*rectenna*) is built and tested for receiving range performance.

1. Introduction

For autonomous powering of sensor nodes in remote or inaccessible areas, wireless power transfer provides the only viable option to power them from an energy source. Due to the low power density of ambient RF at far-field from transmitters, there is a need to optimize each aspect of a wireless RF energy harvester for possible realistic applications. Today remote autonomous sensors are mostly powered by batteries, which have limited lifespan. Renewable powering has the potential to power autonomous sensors perpetually. Due to the expansion of telecommunications technology ambient electromagnetic (EM) power is among the most common sources of ambient energy. There are power transmitters/receivers scattered in practically any society, ranging from television transmission stations to cell phone transmitters and even wireless routers in our homes/offices or mobile phones. These transmitters in our environment and others which are on special dedicated frequencies produce ambient RF power (on the order of microwatts) which can be used as a source for powering remote microwatt budget sensors through wireless energy harvesting. This work presents different matching techniques based on different application requirements using Schottky diode-based RF to DC power converting circuits for wireless remote EM energy harvesting around 434 MHz and 13.6 MHz. Generalized analytical models and limitations of the matched RF to DC power converters are discussed. A wireless RF energy harvester consisting of an antenna and a matched diode rectifier is then realized and its performance tested. Passive wireless energy harvesting also finds applications in near field communications (NFC) [1], RFID tags [2–5], implantable electronics [6,7], and environmental monitoring [8], among others.

1.1. State of the Art

Hertz was the first to demonstrate the propagation of EM waves in free space and to demonstrate other properties of EM waves such as reflection using parabolic reflectors [9]. Wireless power transmission was then investigated and demonstrated for possible wireless remote powering by Tesla. Electromagnetic power beaming for far field wireless power transfer using collimated EM waves was proposed in the 1950s [9]. Recent advances in ultralow power sensors means ambient omni-directional EM power can be used as a source for powering remote sensors without the need to collimate the EM power through the wireless space. Mickle [10] and McSpadden [11] have presented earlier work on wireless energy harvesting systems using Schottky diodes and rectennas where the usability of ambient RF power into DC power was shown. Sample [12] presented a wireless harvester which can harvest EM power from TV and radio base stations transmitting 960 kW of effective radiated power; 60 μW was harvested at a range of about 4 km. Umeda [13] and Le [14] have presented more integrated wireless energy harvesters based on CMOS RF to DC rectifying circuits. CMOS-based rectifying power converters provide full compatibility with standard CMOS technologies and have advantages in batch processes for mass production. The drawback of CMOS-based diode connected transistors is the need to bias the gate of the transistors for the rectifying circuits to effectively function. This gate bias is provided externally, which makes the system not passive. Without the injection of external charges or a biasing of the transistor gate, the circuit has low efficiency, especially when the amplitude of the input voltage is low [15]. Shameli [2] presented a passive CMOS RF to DC power converter with a
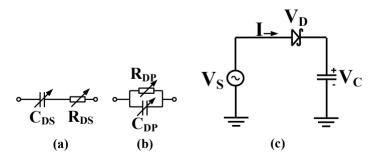
voltage sensitivity of 1 V at −14.1 dBm input, but the circuit efficiency was only 5 %. Zbitou [16] presented an RF to DC converter based on Schottky diodes and achieved 68 % efficiency at 20 dBm RF input power. Ungan [17,18] presented antennas and high quality factor RF to DC power converters at 24 MHz and 300 MHz for RF wireless energy harvesting at -30 dBm input power. The power converter used high quality factor resonators for impedance matching the EM source and the diodes and achieved high open circuit voltage sensitivity of 1 V/μW. Boquete [19] presented a risk assessment system for calculating insurance premiums by monitoring mobile phone usage while driving. This was done by harvesting EM power from detected mobile phone usage during driving for risk assessment. Heikkinen [20] presented rectennas on different substrates at 2.4 GHz using transmisson lines to match the antennas output resistance (*at resonance*) to the rectifying diodes. Akkermans [21] presented a rectenna design by complex conjugating impedance provided by a microstrip structure to a diode so that resonance may be achieved for a working frequency. This design approach may need sophisticated tools to realize and the dominant resonance frequency of the rectenna can be unpredictable in practice. Hagerty [22] presented rectenna arrays for broadband ambient EM harvesting and characterized the harvesters from 2 GHz to 18 GHz; rectennas combine impedance matching the RF rectifying circuit and the antenna into one compact device, but an array of rectennas may increase the overall size of an EM harvester. Herb [23] and Vullers [24] have provided a comprehensive state of the art for micro energy harvesting and have explored the various techniques used for harvesting ambient renewable energy.

2. RF to DC Power Converter

2.1. Diode Rectifier

A junction diode equivalent circuit and simple Schottky diode rectifier are shown in Figure 1. R_{DS} is the diode resultant series resistance, C_{DS} is the diode resultant series capacitance, R_{DP} is the diode resultant parallel resistance, C_{DP} is the diode resultant parallel capacitance, *Vs* is the sinusoidal source voltage and *Vc* is the voltage across the capacitor.

Figure 1. (**a**) Diode series equivalent model, (**b**) Diode parallel equivalent model, (**c**) Simple diode detector.

The diode capacitive impedance is mainly due to the junction capacitances provided by the metal, its passivation and the semiconductor forming the diode. AC power incident on a forward biased diode input is converted to DC power at the output. The current-voltage behavior of a single metal/semiconductor diode is described by the Richardson equation [25] as in Equation (1):

$$I = I_S\left(e^{\left(qV_D/nKT\right)} - 1\right)$$

(1)

where I is the current through the diode, I_S is the saturation current, q is the charge of an electron, V_D is the voltage across the diode, T is the temperature in degrees Kelvin and K is Boltzmann constant. The voltage equation around the loop can be derived from Figure 1(c) and is given in Equation (2):

$$V_D = V_S - V_C$$

(2)

Since the same current flows through the diode and the capacitor, one can find the average current through the circuit by integrating Equation (1) over a time period. By substituting Equation (2) into Equation (1), V_C can be expressed in terms of V_S by averaging the diode current to zero. This is given in Equation (3) [26]:

$$V_C = \frac{KT}{q}\ln\left[\vartheta_0\left(\frac{qV_S}{KT}\right)\right],$$

(3)

where ϑ_0 is the series expansion of the sinusoidal source voltage. Equation (3) can further be simplified for very small amplitude V_S as Equation (4):

$$V_C \approx \frac{qV_S^2}{4KT}$$

(4)

Equation (4) shows that for a small voltage source, the circuit output voltage is proportional to the square of the input sinusoidal voltage; hence it's so-called square law operation. Extensions of this model for voltage multipliers and other input signals are presented in [27] and [28]. Equation (4) further confirms that for low input voltage (power ≤ 10 dBm), an impedance matching network between the source and the diode is necessary to improve the detected output voltage and efficiency.

2.2. Impedance Matching

The maximum power transfer theorem states that the highest power is transferred to the load when the source resistance is the same as the load resistance. For systems with both resistive and reactive impedances from source and load, the source and the load impedance should be adjusted in a way that they are the complex conjugate of each other through impedance matching. For the purposes of this work, a 50 Ω resistive source is chosen as reference for load impedance matching. The antenna which captures the ambient RF signals is tuned to provide this source resistance at resonance for the rectifying circuit in a complete EM wireless remote harvester. The load is the resistance of the Schottky diodes and the actual connected resistance (*remote sensor*). The specific type of matching network which can be used for complex conjugation depends on the nature of load or source impedance, the desired RF to DC converter functionality and other factors like circuit size, cost, *etc.* The response of a matched RF to DC power converter depends on the matching network used as well as the source or load component quality factors and impedances.

2.3. Diode Impedance

Schottky diodes HSMS-285C and HSMS-286C from Avago [29,30] are used to build the RF to DC power converters. The HSMS-285× or 286× series diodes can be operated as zero biased with relatively low forward junction potential. This allows for the realization of completely passive RF to DC power converters for wireless energy harvesting. The HSMS-285C or 286C is a pair of series connected Schottky diodes in a SOT-323 package. The impedance of the HSMS-285C and HSMS-286C diodes was first measured so it can be matched to the resistance (50 Ω) of the antenna source. This is done by connecting the input of the diodes to a network analyzer and measuring the scattering parameters. These scattering parameters are then converted to the corresponding impedances. The input impedance of a diode depends mainly on the resistive and capacitive impedance provided by the junction of the diode and its connected load. For a couple of diodes arranged in a package such as the HSMS-285C or 286C, the input impedance is the vector sum of the impedances provided by each diode in the package arrangement, the extra impedance associated with the packaging and the connected load. The diode measuring board is as shown in Figure 2. The diodes were measured at room temperature for an input power of -30 dBm at a diode connected load of 1 MΩ with a 100 pF filter capacitor. For the sake of this work, the input impedance of the diodes will always be referred to at these connected load conditions.

Figure 2. (**left**) Reference circuit layout for measuring diodes input impedance, (**right**) measuring printed circuit board (PCB) for diodes input impedance on 1 mm FR4 substrate.
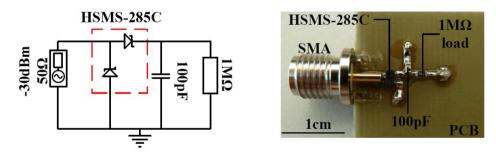
Figure 3. Measured input impedance (Δ resistive, □ capacitive) of HSMS-285C (**left**) and HSMS-286C (**right**) diodes at −30 dBm input with 1 MΩ load and 100 pF filter.
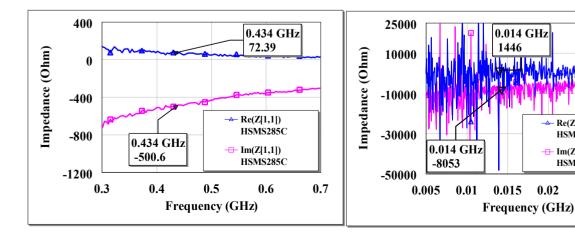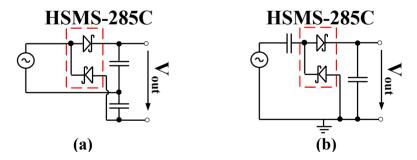
The board is fabricated such that components are soldered directly one into another to prevent additional impedances introduced by copper route. The PCB backside had the ground layer. An example of measured input impedance for HSMS-285C and HSMS-286C is shown in Figure 3.

The diodes quality factor is given by $X_{DS}R_{DS}^{-1}$, where X_{DS} is the resultant series capacitive impedance of the diodes. At an input power of −30 dBm, the measured input impedance of the HSMS-285C diodes is 72–j501 Ω at 434 MHz and 587–j1239 Ω at 13.6 MHz. For HSMS-286C diodes, it is 10–j503 Ω at 434 MHz and ~1.5–j8.1 kΩ at 13.6 MHz for −30 dBm input. The measured impedance of the HSMS-286C diodes at low frequencies (< 60 MHz) shows pronounced fluctuations. The low-frequency excess flicker noise and the shot noise observed in the HSMS-286C have been studied by several authors [31–33]. The pronounced presence of trap states in the depletion region of the semiconductor, mobility fluctuations in carriers, edge effects among other reasons is reported to cause deviations from the ideal Schottky diode behavior and hence generation-recombination noise for some diodes such as the HSMS-286C [34]. When a diode rectifier is matched at a reference operating condition, the matching network may function less effectively at other input power levels, connected load and other operating frequencies. This is due to possible changes in the diode input impedance. Throughout this work the imperfections of the matching circuit at other operating conditions away from the matched reference conditions are accepted without changes to the matching network.

2.4. Voltage Doubler

The Delon voltage doubler and Greinacher doubler are both used to realize the RF to DC power converters presented in this work. The Delon voltage doubler and Greinacher doubler are shown in Figure 4. The diodes output voltage (V_{out}) is doubled what is detected by a simple detector circuit shown in Figure 1. Both doublers produce the same output performance, the only difference is that the Delon doubler has an instantaneous input ground which is not shared with the output.

Figure 4. Circuit diagram of voltage doubler, (**a**) Delon doubler and (**b**) Greinacher doubler.

2.5. Matching Techniques for Antenna Source and RF to DC Power Converter

2.5.1. L-match RF to DC Power Converter

An L-match network converts a source series impedance to its equivalent load parallel impedance or *vice-versa* and tunes out by subtracting or adding any surplus reactance from the load or source with the counter impedance. Series impedance is converted to its parallel equivalent impedance using Equations (5–7):

$$Q_S = \frac{X_S}{R_S} \tag{5}$$

$$Q_P = \frac{R_P}{X_P} \tag{6}$$

where *Xs* is the total series reactive impedance, *Rs* is the total series resistance, R_P is the total parallel resistance, *Xp* is the total parallel reactive impedance, Q_S and *Qp* are the series and parallel quality factors respectively:

$$R_S + jX_S = \frac{R_P \times jX_P}{R_P + jX_P} \tag{7}$$

Equation (7) is the equation of a series sum of impedances and a parallel sum of impedances. It is interesting to note that Q_S and Q_P from an L-matched network may be different from the individual component quality factors as a result of the inherent resistive and reactive impedances in that component. By virtue of Equation (7), Q_S and Q_P must be equal in an L-matched network. Using Equations (5,6) and (7), the ratio of the parallel resistance (*or reactance*) to the series resistance (*or reactance*) can be derived in terms of the quality factors Q_P or Q_S [35]. Since at match conditions, only the resistive impedances dissipate power, the loaded quality factor *Q,* of the L-matched network can be expressed as in Equation (8):

$$R_P = (Q^2 + 1)R_S \tag{8}$$
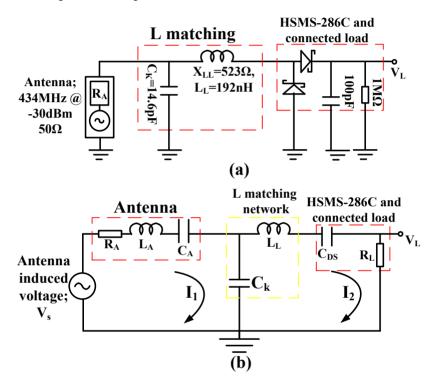
Using Equations (5,6) and (8), series impedance can be converted to its parallel equivalent for a fixed frequency and power level. As an example; a series impedance 72–j501 Ω (HSMS-285C at 434 MHz for −30 dBm input power) is easily converted to −j510(3519)/(−j510 + 3519) Ω as its parallel equivalent with a component quality factor of 6.96. The source resistance is taken as part of the parallel matching network in an L-match circuit if the source series equivalent resistance is greater than the load series equivalent resistance. On the other hand, the load resistance is taken as part of the parallel matching network if the load series equivalent resistance is greater than the source series equivalent resistance. For the purpose of this work, inductors were only used for series impedance matching and capacitors as shunts. This prevents power seeping through any shunt inductor used for impedance matching due the short circuit provided by a shunt inductor to ground and resulting in less output efficiency. Resistors were not used for impedance matching.

2.5.2. L-match RF to DC Converter Generalized Analytical Model

The classical matching technique using Equations (5,6) and (8) is first used to L-match the 50 Ω resistance of the antenna to the resistance of the HSMS-286C diodes (and load) at 434 MHz for −30 dBm input and then the generalized model is discussed. The antenna source resistance was L-matched to the resistance of the diodes (and load). The 50 Ω resistance of the antenna is taken as the parallel matching component and the diodes 10 Ω resistance is the series matching component. The loaded *Q* is found as 2 between the 50 Ω antenna source resistance and the 10 Ω diode series resistance using Equation (8). From this loaded *Q*, a shunt capacitive impedance of 25 Ω (14.6 pF at 434 MHz) using Equation (6) and a series inductive impedance of 20 Ω (7.3 nH at 434 MHz) using Equation (5)

will match the 50 Ω source to the 10 Ω HSMS-286C diodes (and load) series resistance at −30 dBm input. Since the HSMS-286C diodes inherently provides 503 Ω series capacitive impedance at −30 dBm, a resultant series inductive impedance of 523 Ω (192 nH at 434 MHz) is needed to tune the 50 Ω resistive source to the complete HSMS-286C diodes impedance at 434 MHz for −30 dBm input. The L-matched HSMS-286C diodes rectifier is as shown in Figure 5(a).

Figure 5. (**a**) L-match RF to DC harvester using the HSMS-286C diodes at 434 MHz for −30 dBm input. (**b**) Small signal impedance model of a generalized L-matched RF to DC power converter as capacitive coupled series RLC resonator with different resonator elements.

C_K is the tuning capacitance, L_L is the tuning inductance, X_{LL} is the tuning inductive impedance, C_{DS} is the diodes series capacitance, X_{DS} is the diodes series capacitive impedance, V_S is the antenna captured ambient EM voltage, R_A is the resistance of antenna, L_A is the inductance of antenna, C_A is the capacitance of antenna, R_L is the resultant series resistance from the diodes and the connected load resistance, V_L is the resistive load voltage. From Figure 5(a) the power dissipated in the resistance of the diodes (and connected load); P_L is given by Equation (9), where R_L is the series resistance of the diodes and load:

$$P_L = \frac{V_L^2}{R_L} \tag{9}$$

The source power; P_S is given by Equation (10), where V_S* is the root mean squared (RMS) antenna captured source voltage. Half of the source power is transferred to the resistance of the diodes (and connected load) at match conditions as described by the maximum power transfer theorem:

$$P_S = \frac{V_S^2}{R_A} \quad \text{or} \quad P_S = \frac{V_{S*}^2}{2R_A} \tag{10}$$

Equating P_L and half RMS P_S gives a condition of maximum voltage gain for the matched RF to DC power converter shown in Figure 5(a):

$$\frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{\frac{R_L}{R_A}} \tag{11}$$

From Equation (8), substituting the series and parallel resistance ratio into Equation (11) the voltage gain can be expressed in terms of the loaded quality factor as in Equations (12) and (13), where Q is the loaded quality factor of the RF to DC power converter:

$$\frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{\frac{1}{1+Q^2}} \tag{12}$$

Equation (12) is the voltage gain in-terms of the loaded Q if the resistance of the diodes (and connected load) is part of the series matching network and the resistance of the antenna source is part of the parallel matching network as in Figure 5(a). If the resistance of the diodes is part of the parallel matching network, then Equation (13) may be written as the voltage gain in-terms of the loaded Q in an L-matched circuit:

$$\frac{V_L}{V_{S*}} = \frac{1}{2}\sqrt{1+Q^2} \tag{13}$$

Equations (12) and (13) shows that the maximum voltage gain is directly related to the relative differences between the diodes (and connected load) resistance and source resistance at matched conditions or the circuit loaded quality factor. It is interesting to note that the circuit shown in Figure 5(a) has a loaded Q of 2, but an HSMS-286C unloaded quality factor of 50 (at 434 MHz for −30 dBm).

Figure 5(a) is generally modeled as capacitive coupling of two series RLC resonators with a voltage source. This linearized model can be made at any defined frequency and power level. The model however neglects the metal/semiconductor physics of the diode's junction potentials which results in a Schottky barrier. The first series RLC resonator is modeled as impedance from the antenna with or without some passive matching components. The voltage source V_S, is the antenna captured electromagnetic voltage. The second series RLC resonator is the impedance from the diodes (at a defined condition), connected resistance and some passive matching components. Ck is modeled as the coupling element between the two series RLC resonators. Figure 5(b) gives a more general look at the special scenario shown in Figure 5(a). The voltage equations in the two loops are given by Equations (14,15) according to Kirchhoff's voltage loop laws, where ω is the angular frequency and I_1, I_2 are the currents in the first loop and second loop, respectively:

$$V_S = I_1\left[R_A + j\omega L_A - \frac{j}{\omega C_A} - \frac{j}{\omega C_K}\right] + \frac{jI_2}{\omega C_K} \tag{14}$$

$$0 = \frac{jI_1}{\omega C_K} + I_2\left[R_L + j\omega L_L - \frac{j}{\omega C_{DS}} - \frac{j}{\omega C_K}\right] \tag{15}$$

Using Cramers rule, I_2 can be expressed as:

$$I_2 = \frac{\dfrac{-jV_S}{\omega C_K}}{\left[R_A + j\omega L_A - \dfrac{j}{\omega C_A} - \dfrac{j}{\omega C_K}\right]\left[R_L + j\omega L_L - \dfrac{j}{\omega C_{DS}} - \dfrac{j}{\omega C_K}\right] + \dfrac{1}{\omega^2 C_K^2}} \qquad (16)$$

The voltage across R_L is V_L; given by $I_2 R_L$:

$$V_L = \frac{\dfrac{-jV_S}{\omega C_K}R_L}{\left[R_A + j\omega L_A - \dfrac{j}{\omega C_A} - \dfrac{j}{\omega C_K}\right]\left[R_L + j\omega L_L - \dfrac{j}{\omega C_{DS}} - \dfrac{j}{\omega C_K}\right] + \dfrac{1}{\omega^2 C_K^2}} \qquad (17)$$

The voltage gain of the coupled resonator can be expressed as in Equation (18):

$$\frac{V_L}{V_S} = \frac{\dfrac{-jR_L}{\omega C_K}}{\left[R_A + j\omega L_A - \dfrac{j}{\omega C_A} - \dfrac{j}{\omega C_K}\right]\left[R_L + j\omega L_L - \dfrac{j}{\omega C_{DS}} - \dfrac{j}{\omega C_K}\right] + \dfrac{1}{\omega^2 C_K^2}} \qquad (18)$$

At resonance, there is no resultant reactance in the RLC resonators or the capacitive and inductive impedances become equal; hence Equation (19) can be written:

$$\omega L_A - \frac{1}{\omega}\left\{\frac{1}{C_A} + \frac{1}{C_K}\right\} = 0 \text{ and } \omega L_L - \frac{1}{\omega}\left\{\frac{1}{C_{DS}} + \frac{1}{C_K}\right\} = 0 \qquad (19)$$

Equations in Equation (19) can be used to find the resonant frequencies of the series coupled resonator. The voltage gain of the coupled resonator at resonance can then be expressed as in Equation (20):

$$\frac{V_L}{V_S} = V_{gain} = \frac{\dfrac{-jR_L}{\omega C_K}}{R_A R_L + \dfrac{1}{\omega^2 C_K^2}} \qquad (20)$$

where V_{gain} is the voltage gain. V_{gain} at resonance is a function of the resistance of the source and load and the coupling element. The maximum of Equation (20) is obtained when:

$$\frac{dV_{gain}}{dC_K} = 0. \qquad (21)$$

This gives the results as in Equation (22):

$$\frac{dV_{gain}}{dC_K} = \frac{j2R_L}{\omega^3 C_K^4} - j\left\{R_A R_L + \frac{1}{\omega^2 C_K^2}\right\}\frac{R_L}{\omega C_K^2} = 0 \text{ or } R_A R_L^2 = \frac{R_L}{\omega^2 C_K^2} \qquad (22)$$

Equation (22) can be simplified to find $C_{K(max)}$:

$$C_{K(max)} = \pm\frac{1}{\omega}\sqrt{\frac{1}{R_A R_L}} \qquad (23)$$

where C_{Kmax} is the value of the coupling element where maximum power transfer from the first resonator to the second resonator occurs. Using Equations (19) and (23) the unknown optimal matching impedances can be found from the known impedances just like the classical L-matched procedure using Equations (5,6) and (8). By substituting $C_{K(max)}$ into Equation (20) and taking the magnitude of V_{gain}, gives the maximum voltage gain of the coupled series resonator at resonance:

$$\left| \frac{V_L}{V_S} \right| = \frac{1}{2} \sqrt{\frac{R_L}{R_A}} \ \text{ or simply } \ \frac{V_L}{V_{S*}} = \frac{1}{2} \sqrt{\frac{R_L}{R_A}} \tag{24}$$

For wireless harvesters consisting of an antenna and a diode rectifying circuit, the diode resistive impedance at any condition is dependent on the diode realized parameters, signal frequency, connected load and the input power level. The source impedance is determined by the impedance of the antenna. For maximum efficiency, the ratio of the source resistance to the load resistance must tend to zero at matched conditions. The efficiency η of the circuit is given by Equation (25):

$$\eta = \frac{P_L}{P_S} \ ; \eta \rightarrow 1 \ \ when \ \frac{R_A}{R_L} \rightarrow 0 \tag{25}$$

2.5.3. L-Match RF to DC Converter Experimental Results and Discussion

The presented circuit was L-matched between the 50 Ω resistance of the antenna source and the resistance of the HSMS-285C diodes (and load) at 434 MHz for −30 dBm input as shown in Figure 6. Since the series equivalent resistance of the HSMS-285C diodes and load (72 Ω) is greater than the 50 Ω series resistive antenna source, the diode is taken as parallel matching network with a parallel equivalent impedance of $-j510(3519)/(-j510 + 3519)$ Ω. The analysis follows the same procedure as in Section 2.5.2 after this step. Figure 6(b) shows the resultant L-matched RF to DC power converter. $C_{DP}{*}$ is the resultant shunt matching capacitance.

Figure 6. (**a**) L-matched impedance circuit for matching the HSMS-285C diodes at 434 MHz for −30 dBm input. (**b**) Resultant network, (**c**) PCB layout of the L-matched Delon doubler with adjusted values on FR4 substrate (**d**) Fabricated PCB of the L-network matched Delon voltage doubler.

Figure 6. *Cont.*

(c)

(d)
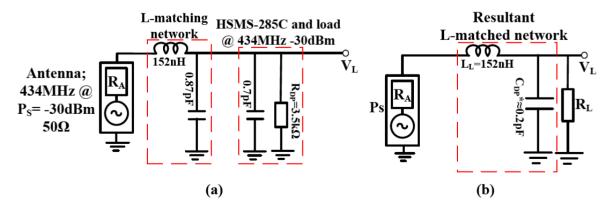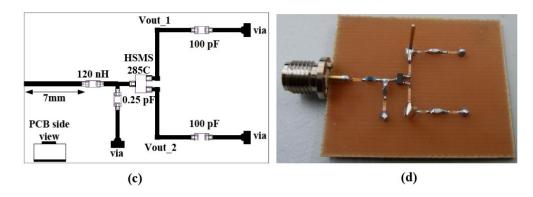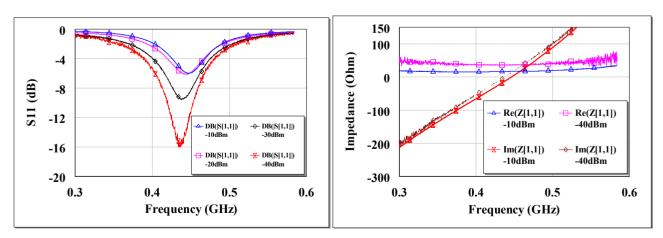
Figure 6(a,b) assume perfect characteristic impedance between the various components in the matched circuit. When a copper route is introduced between components and on a material substrate, it must be accounted for in the total impedance as seen by the source or load. This PCB impedance compensation is carried out in Advance Design Systems (ADS) from Agilent [36]. ADS has extensive models for microstrip substrates to account for its impedances. The optimized layout using ADS microstrip models and its compensated values in the passive tuning components for a Delon doubler is shown in Figure 6(c).

The circuit reflection coefficient (S_{11}) and input impedance at open circuit are shown in Figure 7. There is high return loss and resonance around 434 MHz. The circuit input impedance at open circuit conditions is ~38 Ω at resonance for −40 dBm and ~17 Ω at −10 dBm input.

The measured L-matched circuit efficiency and voltage sensitivity is shown in Figure 8. The maximum measured L-matched efficiency at −30 dBm is 22% at ~20 kΩ load and an open circuit voltage of 124 mV. At −10 dBm, the maximum efficiency and open circuit voltage is 47% and 2 V respectively. At the optimal load of ~20 kΩ, the detected voltage is 58 mV and 1 V at −30 dBm and −10 dBm respectively.
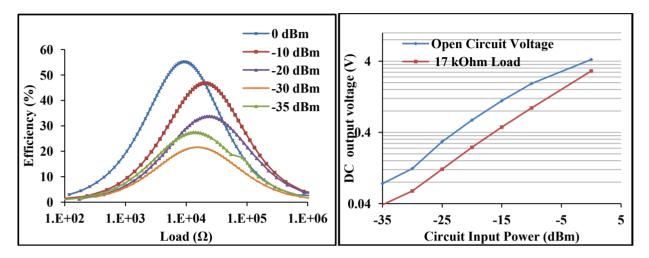
Figure 7. Measured open circuit S_{11} of the L-matched Delon circuit at different input power levels from a 50 Ω source (**left**), measured open circuit input impedance at −10 dBm and −40 dBm of the L-matched circuit (**right**).

The open circuit voltage gain is 25 at −30 dBm and 40 at −10 dBm. The maximum measured efficiency at −35 dBm is 27%. This is higher than that of −30 dBm due to the better matched circuit

impedance at −35 dBm (35 Ω) than at −30 dBm (27 Ω) input. The L-matched RF to DC power converter has a loaded *Q*, sensitivity and efficiency determined mainly by the diodes resistance, diodes junction potential, connected resistance and antenna source resistance at matched conditions.

Figure 8. Measured L-matched circuit efficiency *versus* resistive load at various input power levels at 434 MHz (**left**), measured open circuit voltage and at 17 kΩ load *versus* input power at 434 MHz (**right**).

2.5.4. PI-match RF to DC Power Converter

A highly selective or small frequency bandwidth RF power converter is realized with a PI-network in-between the source impedance from the antenna and the diode rectifier. A PI-network is a 'back to back' L-network that are both configured to match the load and source impedance to an invisible resistance located at the junction between the two L-networks [37]. The quality factor of the L-network with the parallel resistance is given by Equation (26):

$$Q_P^* = \sqrt{\frac{R_P}{R^*} - 1},\qquad(26)$$

where R_P is the parallel resistance, R^* is a virtual resistance and Q_P^* is the quality factor of the L-network with the parallel resistance. The quality factor of the L-network with the series resistance is given by Equation (27):

$$Q_S^* = \sqrt{\frac{R_S}{R^*} - 1},\qquad(27)$$

where Q_S^* is the quality factor of the L-network with the series resistance. The unloaded quality factor; Q_S^* or Q_P^* is set higher than what is normally achieved with a single L-network [37] to realize the small frequency bandwidth circuit. The resistance of the load is assigned the parallel network in a PI-matched circuit if its series equivalent resistance is higher than the source series equivalent resistance; the opposite is true if the source is higher than the load. Equation (26) and Equation (27) are synonymous to Equation (8), except the lowest resistive impedance in Equation (8) is substituted with the virtual resistance which is dependent on the newly desired circuit selectivity. From Equations (26) and (27) the loaded quality factor of the PI-matched circuit can be written as Equation (34) in terms of Q_S^* and Q_P^*:

$$Q^2 = \left[\left(\frac{Q_P^{*2} + 1}{Q_S^{*2} + 1} \right) - 1 \right], \tag{28}$$

where Q is the loaded quality factor of the PI-network. Q_S^* or Q_P^* are the unloaded quality factors of the PI-matched network. The larger value among the unloaded quality factors result in small frequency bandwidth response which is desired when matching a source and load impedance with a PI-network. Some authors approximate the highest value of Q_S^* or Q_P^* or their algebraic sum as the loaded quality factor of the PI-network as in [35] and [37], but Equation (28) gives the exact loaded Q of the PI-matched circuit in terms Q_S^* and Q_P^*. This allows for the correct estimation of the maximum voltage gain from the loaded quality factor.

2.5.5. Selectivity RF to DC Converter Generalized Analytical Model

An example of a PI-matched RF to DC converter using the HSMS-285C diodes operating at 434 MHz for −30 dBm input is presented first and then the generalized model is discussed. The circuit is matched for Q_P^* of 60 between the antenna and the resistance of the diodes as shown in Figure 9.

Figure 9. Impedance diagram of PI-matched RF power converter; (**a**) Impedance diagram of 50 Ω source and the HSMS-285C diodes at 434 MHz, (**b**) Resultant PI matched network between the antenna source and load resistance.

(**a**)   (**b**)
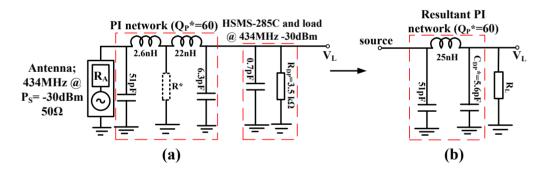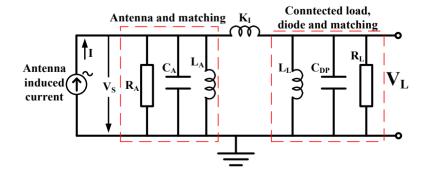
Figure 9(b) can also be modeled as an inductive coupling of two parallel RC circuits. A more general look at such a circuit is shown in Figure 10, as an inductive coupling of two parallel RLC resonators with a current source.

Figure 10. Inductive coupled parallel RLC small signal model of a generalized PI-matched antenna and diode rectifier.

The first parallel RLC resonator is modeled as impedance from the antenna and some passive matching components. The second parallel RLC resonator is modeled as impedance from the linearized diodes, its connected load and some passive matching components. I is the antenna induced current, V_S this time is the voltage across the parallel R_A and K_1 is the coupling element between the two parallel resonators. Using Kirchoff's current laws, the node equations can be expressed as Equations (29) and (30):

$$I = V_S\left[\frac{1}{R_A} + j\omega C_A - \frac{j}{\omega L_A} - \frac{j}{\omega K_1}\right] + \frac{jV_L}{\omega K_1} \tag{29}$$

$$0 = \frac{jV_S}{\omega K_1} + V_L\left[\frac{1}{R_L} + j\omega C_{DP} - \frac{j}{\omega L_L} - \frac{j}{\omega K_1}\right] \tag{30}$$

Load voltage (V_L) and the source voltage (V_S) at resonance are given by the equations in Equation (31). The resonance frequencies are given by Equation (32):

$$V_L = \frac{\dfrac{-jI}{\omega K_1}}{\left[\dfrac{1}{R_A R_L} + \dfrac{1}{\omega^2 K_1^2}\right]} \quad \text{and} \quad V_S = \frac{\dfrac{I}{R_L}}{\left[\dfrac{1}{R_A R_L} + \dfrac{1}{\omega^2 K_1^2}\right]} \tag{31}$$

$$\omega C_A - \frac{1}{\omega}\left\{\frac{1}{L_A} + \frac{1}{K_1}\right\} = 0 \text{ and } \omega C_{DP} - \frac{1}{\omega}\left\{\frac{1}{L_L} + \frac{1}{K_1}\right\} = 0 \tag{32}$$

From V_L and Vs in Equation (31), the voltage gain at resonance can be expressed as:

$$\frac{V_L}{V_S} = \frac{R_L}{j\omega K_1} \tag{33}$$

The maximum of Equation (33) is obtained when:

$$j\omega K_1 \to 0 \quad or \quad R_L \to \infty \tag{34}$$

Since $j\omega K_1$ is restricted by the conditions in Equation (32) to attain resonance, one cannot manipulate $j\omega K_1$ alone without changing the resonance conditions. What can drive the voltage gain is if R_L is very large at resonance conditions. If the input impedance (V_S/I) of the coupled resonator is maximum at resonance, conditions in Equation (35) hold:

$$\left(\frac{Vs}{I}\right) \to maximum \quad when \quad \frac{R_L}{\omega^2 K_1^2} \to 0 \tag{35}$$

Equation (36) may be assumed when $\dfrac{R_L}{\omega^2 K_1^2} \to 0$:

$$\frac{V_S}{I} = R_A \tag{36}$$

Under these conditions and an optimal coupling coefficient K_{1max}, the maximum voltage gain of the parallel coupled resonator can be written as in Equation (37), where K_{1max} is given by Equation (38):

$$\left|\frac{V_L}{V_S}\right| = \left|V_{gain}\right| = \left|\frac{1}{2}\sqrt{\frac{R_L}{R_A}}\right| \tag{37}$$

$$K_{1(\text{max})} = \mp\frac{1}{\omega}\sqrt{R_A R_L} \tag{38}$$

The analysis of Section 2.5.2 and parallel coupled RLC resonators show that any antenna and matched rectifying diode can be described as an equivalent circuit of a coupled resonator at a defined operating point. This general model can be applied to optimize other harvesters with complex output impedance such as piezo-harvesters or vibration harvesters for maximum transfer of power or voltage to its connected load. The model can also be applied to near field magnetically coupled antennas/coils for optimization.

2.5.6. Broadband RF to DC power converter

A broadband network is preferred when an RF to DC power converter is to be operated for a wide range of frequencies. A broadband converter is realized by connecting successive L-networks together in a multi-network between the antenna source and the rectifying diodes. The result is broadband or multiband RF power converter around certain frequencies. This can be deduced from the general model of a coupled resonators that by choosing certain passive components between a source and the load, it is possible to have more frequencies (ω) fulfilling Equation (32) and hence a result of multiple resonant frequencies or broader bandwidth at match conditions. For a two stage L-connected match, the quality factor of the L-network with the parallel resistance is given by Equation (39):

$$Q_P^* = \sqrt{\frac{R_P}{R^*} - 1} \tag{39}$$

The quality factor of the L-network with the series resistance is given by Equation (40):

$$Q_S^* = \sqrt{\frac{R_*}{R_S} - 1} \tag{40}$$

From Equations (39) and (40) the loaded quality factor of the two stage L-connected broadband network may be written as Equation (41) in terms of the unloaded quality factors; Q_S^* and Q_P^*:

$$Q^2 = \{(Q_P^{*2} + 1)(Q_S^{*2} + 1)\} - 1 \tag{41}$$

R^* in this case may be chosen if it is larger than R_S and lower than the R_P. The highest possible bandwidth between a resistive source and resistive load is found for a virtual resistance (R^*) given in Equation (42) [37]:

$$R^* = \sqrt{R_S R_P} \tag{42}$$

For complex loads such as rectifying diodes or transistors, the largest achievable bandwidth prescribed by Equation (42) is limited by the load or source component quality factor, since Equation (42) does not take into account reactive impedance associated with the source or load.

2.5.7. Broadband-Match RF to DC Converter Results and Discussion

The antenna source resistance was broadband matched to the HSMS-285C diodes (and load) resistance at −30 dBm input around 434 MHz. For a desired Q_P* and Q_S* of 2.7 there is ~0.4 pF inherent diode capacitance which is un-tuned using a two stage L-matching network [Figure 11(b)]. This is because the HSMS-285C diodes provides an inherent component quality factor of 6.96 at 434 MHz for −30 dBm input, hence a broadband circuit with Q_P* lower than this inherent component quality factor of the diodes (and load) is difficult to achieve without trade-offs. However, connected L-networks with Q_P* as high as the diode component quality factor may perform worse than a single L-matched network with similar loaded quality factor. This is due to redundant components of the connected L-networks which have inherent losses.

Figure 11. Impedance diagram of broadband RF power converter; (**a**) Broadband match around 434 MHz with loaded Q of 2.7, (**b**) Resultant impedance matching network with un-turned capacitance of 0.4 pF.

(a)   **(b)**
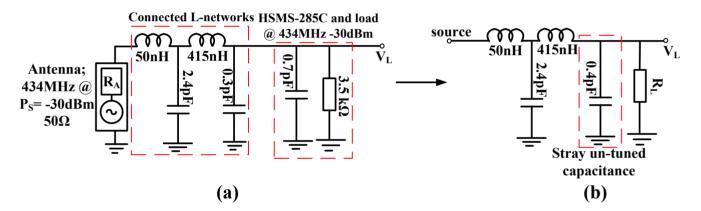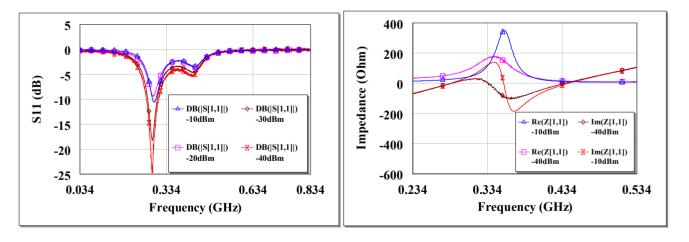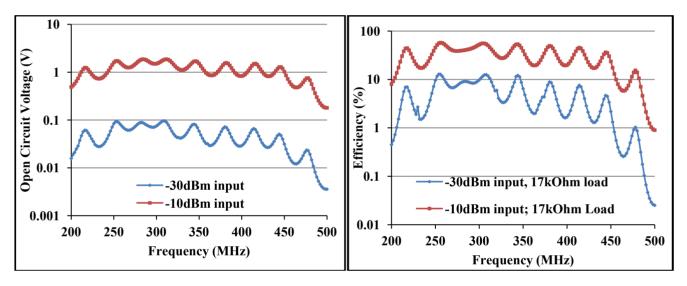
Figure 12. Measured open circuit S_{11} of the broadband circuit around 434 MHz at different input power levels from a 50 Ω source (**left**), measured open circuit input impedance at −10 dBm and −40 dBm of the broadband circuit (**right**).

Therefore the broadband circuit is matched for Q_P* of 2.7, notwithstanding the un-tuned shunt capacitance as can be seen in Figure 11(b). Figure 12 shows the circuit S_{11} at various input power levels and input impedance at open circuit conditions. From Figure 12 (left) there is ~−5 dB return loss

from 200 MHz to 500 MHz providing an operating band of ~300 MHz. The impedance of the circuit shows resonances at ~290 MHz and ~450 MHz as shown in Figure 12(right). A third resonance occurs around 356 MHz at −10 dBm as the frequency of highest harvester input resistance (~350 Ω) and where the reactive impedances approach their extremes. Figure 12 show that a wireless EM harvester can exhibit different resonance scenarios depending on the dominant instantaneous conditions. The efficiency and voltage sensitivity of the broadband matched wireless EM harvester are shown in Figure 13. The average open circuit voltage is 47 mV and 1.1 V at −30 dBm and −10 dBm, respectively, when operating from 200 MHz to 500 MHz.

Figure 13. Measured open circuit voltage *versus* frequency sweep from 200 MHz to 500 MHz for −10 dBm and −30 dBm (**left**), measured efficiency at 17 kΩ load *versus* frequency sweep for −10 dBm and −30 dBm (**right**).

The broadband circuit achieves average efficiency of 5% at 17 kΩ load for −30 dBm and 30% at 17 kΩ load for −10 dBm input power from 200 MHz to 500 MHz. Figure 13 further confirm a direct link between frequency response and the unloaded quality factors. For Q_S* and Q_P* of ~2.7, the circuit response is broadband around 434 MHz.

2.6. High Voltage Sensitive RF to DC Converter
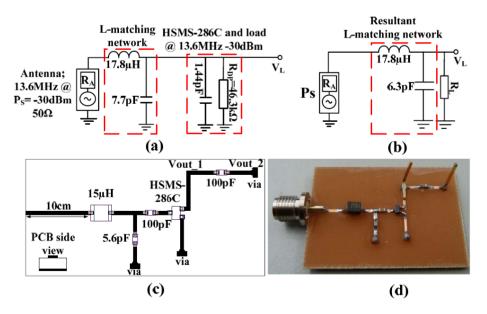
The current state of the art low power remote sensors would require a DC voltage supply of about 1 V and DC current of about 30 μA for operation. Therefore, the issue is not only how efficient a wireless EM harvester is in converting RF to DC power, but also what the output DC voltage and current of the EM harvester are at the RF input power level [38]. Equations (11,24) and (33) show that the maximum voltage sensitivity of a coupled resonator system or an RF to DC power converter is mostly related to the load and the source resistances at resonance. Therefore high voltage sensitive wireless EM harvester can be achieved with a diode voltage doubler with a very high input resistance relative to the antenna source without the need to cascade the diodes as in voltage multipliers. If the diodes been used for the RF to DC power conversion cannot provide high resistive impedance at the working frequency relative to the antenna source, then a DC-DC converter can be applied after the EM harvester as presented in [39] or the diodes may be cascaded by way of multipliers as presented in our

earlier work [40] and by several other authors [3,5,14]. In case of multipliers, the input voltage ought to be high enough to overcome the junction potential of the several diodes in the multiplier network. If frequency is not a constraint, then a frequency sweep *versus* impedance for the diodes can be made and the frequency where the diodes exhibits high resistive impedance can be used to realize high voltage sensitive wireless RF harvester. For Schottky diodes, high resistive impedance occurs mostly at lower frequencies (see Figure 3). The measured voltage gain of a high resistive diode pair (voltage doubler) is presented in the next results.

2.6.1. High Voltage Sensitive RF to DC Converter Results and Discussion

The presented result was L-matched using 50 Ω resistance of the antenna source and the resistance of the HSMS-286C diodes (and load). The HSMS-286C diodes do provide high resistive impedance at low frequencies; notwithstanding the flicker noise which causes its resistive (and reactive) impedance to fluctuate. The HSMS-286C has low forward junction potential (~350 mV at 1 mA) per diode and series impedance of ~1.5–j8.1 kΩ or parallel impedance of ~−j8.3(46.3)/(−j8.3 + 46.3) kΩ at 13.6 MHz for −30 dBm input. Even though the HSMS-286C diodes unloaded component quality factor at 13.6 MHz is similar to that of the HSMS-285C diodes at 434 MHz, the elevated resistive impedance at 13.6 MHz fulfills the condition for high voltage sensitivity relative to a 50 Ω antenna source at resonance conditions.
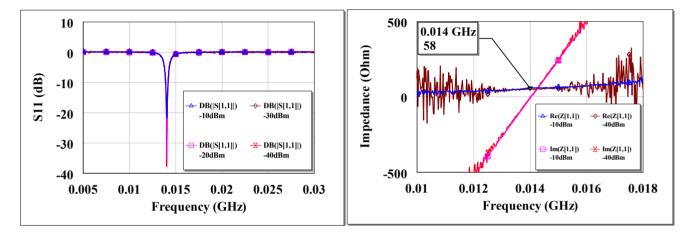
Figure 14. (**a**) L-matched impedance diagram for matching the HSMS-286C diodes at 13.6 MHz at -30 dBm input. (**b**) Resultant network, (**c**) PCB layout of the L-matched Greinacher doubler with adjusted values due to impedances provided by copper route on FR4 substrate with thickness of 1 mm. (**d**) Fabricated PCB of the L-matched RF to DC power converter.

The high voltage sensitive EM harvester operating at 13.6 MHz is as shown in Figure 14. On the realized PCB is a Greinacher doubler. An inductance of 15 µH and a shunt capacitance of 5.6 pF were the adjusted values after the microstrip contributions.
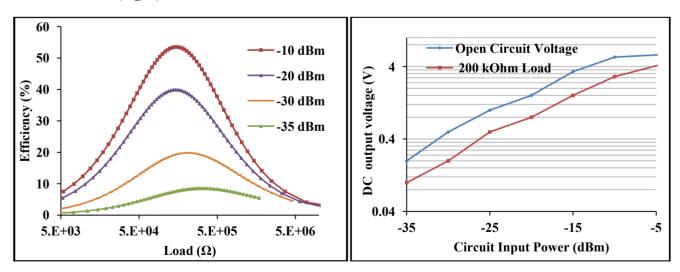
The measured S_{11} and input impedance at open circuit are shown in Figure 15. There is high return loss and resonance around 13.6 MHz. The circuit input impedance at open circuit conditions is 58 Ω at resonance for both −40 dBm and −10 dBm.

Figure 15. Measured open circuit S_{11} of the L-matched HSMS-286C diodes at 13.6 MHz for different input power levels from a 50 Ω source (**left**), measured open circuit input impedance at −10 dBm and −40 dBm of the L-matched HSMS-286C diode at 13.6 MHz (**right**).

The efficiency and voltage sensitivity of the high voltage sensitive wireless EM harvester are shown in Figure 16.

Figure 16. Measured circuit efficiency *versus* load at various input power levels at 13.6 MHz (**left**), measured open circuit voltage and at 200 kΩ load *versus* input power at 13.6 MHz (**right**).
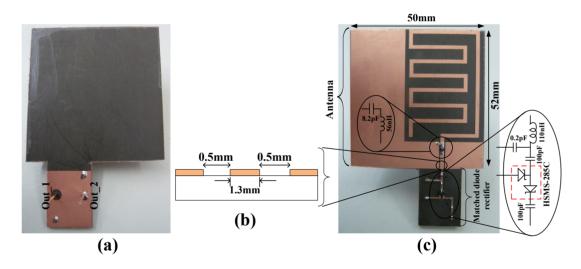
The maximum measured efficiency at −30 dBm is 20% for ~200 kΩ load and an open circuit voltage of 0.5 V. At −10 dBm, the maximum efficiency and open circuit voltage are 54% and 5.4 V respectively. At the optimal load of ~200 kΩ, the detected voltage is 0.2 V and 2.9 V at −30 dBm and −10 dBm respectively. The open circuit voltage gain is 100 at −30 dBm and 108 at −10 dBm.

Even though the RF to DC converter presented in Section 2.5.3 is the same as the L-match circuit realized with the HMSM-286C diodes at 13.6 MHz, the voltage gain is increased by a factor of 4 due to the large difference between the diodes (and load) resistance and source resistance so that at matched conditions high voltage gain occurs. The loaded Q of the L-matched circuit is 30 which results in small frequency bandwidth just like a PI-matched diode rectifier presented in our earlier work [40]. From this result and the results from our earlier presented PI-matched EM harvester, it can be inferred that all high loaded Q RF to DC circuits have high selectivity but not all highly selective RF to DC circuits have high loaded Q. The voltage sensitivity of the matched HSMS-286C diode at 13.6 MHz can be improved if its resistive impedance is not lowered by the flicker noise.

3. Wireless EM Power Harvester

A wireless EM harvester, consisting of a rectifying antenna (*rectenna*) was designed to find a compromise between size and performance of its antenna. The rectenna is shown in Figure 17.
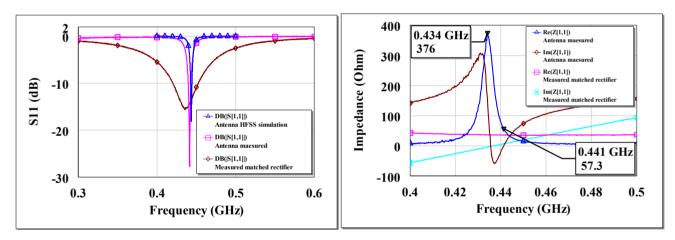
Figure 17. Rectenna realized on a Duroid 5880, 1.57 mm substrate. (**a**) Backside of the rectenna, (**b**) cross-section of antenna output coplanar stripline dimensions (**c**) frontside of the rectenna.

The antenna (planar) part of the rectenna is based on our earlier work [41]. In contrast to the earlier presented antenna, this rectenna is realized on a Duroid [42] substrate of thickness 1.57 mm. Duroid 5880 has lower loss tangent of 0.0004 at 1 MHz compared to 0.025 at 1 MHz for FR4. This means there is less loss in the transmission of signals on a Duroid PCB at this frequency range. The antenna part is fabricated to resonate around 434 MHz; hence its dimensions of 5 × 5.2 cm make it electrically small. The antenna is tuned with a chip inductor and a capacitor to achieve the resonance frequency around 434 MHz [Figure 17(c)]. This is done at a cost of reduced antenna radiation efficiency. An antenna is one of the few components the size of which is related to the operating frequency. Thus, if the size of an antenna is fixed, resonance frequency reduction of the antenna can only be achieved with penalty factors [10]. The antenna's output impedance is tuned with the dimensions of the coplanar stripline as shown in Figure 17(b).

HFSS [43] was used to simulate the presented antenna and to find the correct capacitive and inductive components for frequency tuning before the optimized design was fabricated. The simulated antenna resonances occur at 438 MHz and 445 MHz. At these frequencies, the radiation efficiency is 20% and a peak gain of −6 dBi. The rectifying part of the rectenna consists of L-matched HSMS-285C diodes (Figure 17(c)). The L-matched HSMS-285C part of the rectenna can be engineered to be as small as possible if required. The separate parts of the rectenna were characterized by terminating their ends and measuring the individual reflection coefficients just like the power converters presented in Section 2. Figure 18 shows the measured antenna and matched rectifier individual S_{11} and impedance. Figure 18 (left) also show the HFSS simulated S_{11} results. From Figure 18 (right), the measured antenna resonance where the input impedance is at maximum is ~434 MHz. At ~434 MHz, the antenna input resistance is 376 Ω and the reactive impedances approach their extreme (*so called anti-resonance*). The other resonance occurs when the input resistance is 'finite' and the reactive impedance is zero; at ~441 MHz. The input resistance is 57 Ω at ~441 MHz. The rectifier circuit is matched for the antenna's resistance at ~441 MHz.

Figure 18. Antenna HFSS simulated, antenna measured, and measured L-matched diode rectifier S_{11} on a Duroid 5880 PCB for −30 dBm input (**left**), Measured open circuit input impedance of antenna and rectifier at −30 dBm input (**right**).
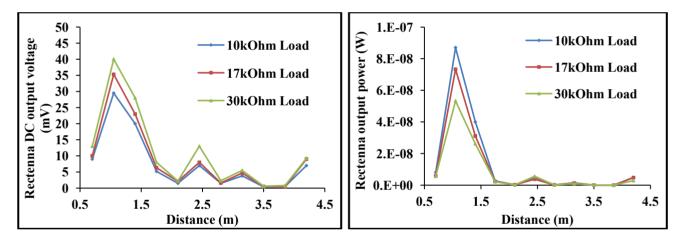
3.1. EM Range Results and Discussion

At far field between wireless EM transmitting and receiving antenna, the coupling mechanism between the transmitting and receiving antenna is neither capacitive nor inductive as is the case for the RF to DC converters. The coupling is radiative which can be described by the Friis equation of transmission on the assumption that the transmitting and receiving antenna are in free space [44]. A modified Friis equation for a transmitting and receiving antenna at far-field (R >> λ and R >> transmitting antenna largest dimension) to each other at a specified direction is given by Equation (43) [45]. Equation (43) assumes real world open space conditions:

$$\frac{P_r}{P_t} = F_{envt} G_t G_r \left(\frac{\lambda}{4\pi R} \right)^2, \tag{43}$$

where Pr is the power at the receiving antenna port, Pt is the power supplied at the transmitting antenna port, F_{envt} is a factor accounting for environmental effects as such ground reflections among others, Gt and Gr are the transmitting and receiving antenna gain (*at specified direction*) respectively. R is the distance between the transmitting and receiving antenna and λ is the wavelength of the transmitting EM wave. The rectenna receiving range measurements were carried out in an open space (*hall*) with the antennas 2 m above ground level. The transmitting and receiving antennas were arranged in the direction of their peak gain. The rectenna range performance is shown in Figure 19. According to Equation (43), the efficiency of RF power transferred between a sending and receiving antenna depends on controllable factors like the gain of the antennas in the arranged direction and the radiation efficiency of the antennas. Since the receiving/transmitting antenna's incorporated in remote harvesters for sensor powering are normally small in relation to their operating frequencies, they tend to be less efficient.

Figure 19. Rectenna receiving range performance by sending 17 dBm (50 mW) at a gain of −6 dBi at 437 MHz. Output DC voltage *versus* receiving distance for different loads (**left**), loads output power *versus* receiving distance (**right**).

The efficiency of the rectenna's antenna is ~20% at resonance. A 'perfectly' matched RF to DC power converter operating in its square law region has efficiencies in the region of 20% as depicted in Section 2. The transmitting antenna was the same as the antenna incorporated in the rectenna. By transmitting the EM power with a small antenna (5 cm × 5.2 cm) at 437 MHz with efficiency of ~20% and at a gain of −6 dBi, the power delivered by the rectenna is generally low at far-field from the transmitter as can be seen in Figure 19. A mediocre transmitting antenna was used to transmit the EM waves due to limitations in the European Union about transmitting EM power at certain frequencies; so the goal in the rectenna range experiment is to show the lowest limit functionality of such a harvester. At 4.2 m from the electrically small transmitting antenna transmitting at 17 dBm, the rectenna harvested DC voltage and power are 9 mV and 5 nW respectively for 10 kΩ load. It can be seen from Figure 19 that the harvested voltage/power generally degrades as an inverse square of distance from transmitter as described by Friis equation. The measured received power however alternate along this R^{-2} fit as shown in Figure 19. This anomaly is accounted for by F_{envt} [Equation (43)] as influence of ground reflections and polarization in real world open field measurements [45]. For any particular distance R, the signals reflected from ground can be constructive with the direct signal to the rectenna,

in which case the measured power may be higher than that predicted by the original Friss equation as in [44]. The ground effect can also be destructive, in which case the measured power will be lower than what is predicted by the original Friis equation.

4. Conclusions

Optimization of Schottky diode-based RF to DC power converters using different matching techniques for wireless EM energy harvesting applications is presented. Using scattering parameters for small signal modeling, it is shown that wireless EM harvesters can be generally described as coupled resonators with efficiencies and maximum voltage sensitivity depending mostly on the source and load resistances under matched conditions. The analytical models allow systematic control in the design of passive wireless EM harvesters. Based on these analyses, a rectenna is built and tested for lower limit functionality from harvesting ambient EM waves. The analysis presented in this work may also be applied to optimize derivatives of wireless EM harvesters like RFID tags, NFC, wireless chargers *etc.*, for efficient powering of their sensors or integrated circuits. Generally, most energy harvesters and their matched loads can be described as coupled resonators and thus may be optimized with the methods presented in this work.

Acknowledgments

References

1. Fischer, J. NFC in cell phones: The new paradigm for an interactive world [Near-Field Communications]. *IEEE Commun. Mag.* **2009**, *47*, 22–28.
2. Shameli, A.; Safarian, A.; Rofougaran, A.; Rofougaran, M.; de Flaviis, F. Power harvester design for passive UHF RFID tag using a voltage boosting technique. *IEEE Trans. Microw. Theor. Tech.* **2007**, *55*, 1089–1097.
3. Karthaus, U.; Fischer, M. Fully integrated passive UHF RFID transponder IC with 16.7-μW minimum RF input power. *IEEE J. Solid-State Circuits* **2003**, *38*, 1602–1608.
4. Lee, S.H.; Jin, I.S. Interoperation of an UHF RFID reader and a TCP/IP device via wired and wireless links. *Sensors* **2011**, *11*, 10664–10674.
5. Mandal, S.; Sarpeshkar, R. Low-power CMOS rectifier design for RFID applications. *IEEE Trans. Circuits Syst. I: Regul. Papers* **2007**, *54*, 1177–1188.
6. Yu, H.; Bashirullah, R. A Low Power ASK Clock and Data Recovery Circuit for Wireless Implantable Electronics. In *Proceedings of IEEE Custom Integrated Circuits Conference, CICC '06*, San Jose, CA, USA, September 2006; pp. 249–252.
7. Hannan, M.A.; Abbas, S.M.; Samad, S.A.; Hussain, A. Modulation techniques for biomedical implanted devices and their challenges. *Sensors* **2011**, *11*, 297–319.

8. Cho, N.; Song, S.-J.; Lee, J.Y.; Kim, S.; Kim, S.; Yoo, H.-J. A 8-μW, 0.3-mm2 RF-powered Transponder with Temperature Sensor for Wireless Environmental Monitoring. In *Proceedings of IEEE International Symposium on Circuits and Systems, ISCAS*, Kobe, Japan, May 2005; Volume 5, pp. 4763–4766.

9. Brown, W.C. The history of power transmission by radio waves. *IEEE Trans. Microw. Theor. Tech.* **1984**, *32*, 1230–1242.

10. Mickle, M.H.; Mi, M.; Mats, L.; Capelli, C.; Swift, H. Powering autonomous cubic-millimeter devices. *IEEE Antennas Propag. Mag.* **2006**, *48*, 11–21.

11. McSpadden, J.O.; Chang, K. A Dual Polarized Circular Patch Rectifying Antenna at 2.45 GHz for Microwave Power Conversion and Detection. In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, San Diego, CA, USA, May 1994; Volume 3, pp. 1749–1752.

12. Sample, A.; Smith, J.R. Experimental Results with Two Wireless Power Transfer Systems. In *Proceedings of the 4th International Conference on Radio and Wireless Symposium*, San Diego, CA, USA, January 2009; pp. 16–18.

13. Umeda, T.; Yoshida, H.; Sekine, S.; Fujita, Y.; Suzuki, T.; Otaka, S. A 950-MHz rectifier circuit for sensor network tags with 10-m distance. *IEEE J. Solid-State Circuits* **2006**, *41*, 35–41.

14. Le, T.; Mayaram, K.; Fiez, T. Efficient far-field radio frequency energy harvesting for passively powered sensor networks. *IEEE J. Solid-State Circuits* **2008**, *43*, 1287–1302.

15. Yuan, F. *CMOS Circuits for Passive Wireless Microsystems*; Springer: New York, NY, USA, 2011.

16. Zbitou, J.; Latrach, M.; Toutain, S. Hybrid rectenna and monolithic integrated zero-bias microwave rectifier. *IEEE Trans. Microw. Theor. Tech.* **2006**, *54*, 147–152.

17. Ungan, T.; Le Polozec, X.; Walker, W.; Reindl, L. RF Energy Harvesting Design Using High Q Resonators. In *Proceedings of IEEE MTT-S International Microwave Workshop on Wireless Sensing, Local Positioning and RFID, IMWS 2009*, Cavtat, Croatia, September 2009; pp. 1–4.

18. Ungan, T.; Freunek, M.; Muller, M.; Walker, W.D.; Reindl, L.M. Wireless Energy Transmission Using Electrically Small Antennas. In *Proceedings of IEEE Radio and Wireless Symposium, RWS '09*, San Diego, CA, USA, January 2009; pp. 526–529.

19. Boquete, L.; Rodríguez-Ascariz, J.M.; Barea, R.; Cantos, J.; Miguel-Jiménez, J.M.; Ortega, S.; Data acquisition, analysis and transmission platform for a pay-as-you-drive system. *Sensors* **2010**, *10*, 5395–5408.

20. Heikkinen, J.; Salonen, P.; Kivikoski, M. Planar Rectennas for 2.45 GHz Wireless Power Transfer. In *Proceedings of Radio and Wireless Conference, RAWCON 2000*, Denver, CO, USA, 10–13 September 2000; pp. 63–66.

21. Akkermans, J.A.G.; van Beurden, M.C.; Doodeman, G.J.N.; Visser, H.J. Analytical models for low-power rectenna design. *IEEE Antenn. Wireless Propag. Lett.* **2005**, *4*, 187–190.

22. Hagerty, J.A.; Helmbrecht, F.B.; McCalpin, W.H.; Zane, R.; Popovic, Z.B. Recycling ambient microwave energy with broad-band rectenna arrays. *IEEE Trans. Microw. Theor. Tech.* **2004**, *52*, 1014–1024.

23. Harb, A. Energy harvesting: State-of-the-art. *Renew. Energ.* **2011**, *36*, 2641–2654.

24. Vullers, R.J.M.; van Schaijk, R.; Doms, I.; van Hoof, C.; Mertens, R. Micropower energy harvesting. *Solid-State Electron.* **2009**, *53*, 684–693.

25. Sah, C.-T. *Fundamentals of Solid-State Electronics*; World Scientific: Singapore, Singapore, 1991.

26. Wetenkamp, S. Comparison of Single Diode *vs.* Dual Diode Detectors for Microwave Power Detection. In *Proceedings of IEEE MTT-S International Microwave Symposium Digest*, Boston, MA, USA, June 1983; pp. 361–363.

27. Cardoso, A.J.; Schneider, M.C.; Montoro, C.G. Design of Very Low Voltage CMOS Rectifier Circuits. In *Proceedings of Circuits and Systems for Medical and Environmental Applications Workshop (CASME)*, Merida, Mexico, December 2010; pp. 1–4.

28. Cardoso, A.J.; de Carli, L.G.; Galup-Montoro, C.; Schneider, M.C. Analysis of the Rectifier Circuit Valid Down to Its Low-Voltage Limit. *IEEE Trans. Circuits Syst. I: Regul. Papers* **2012**, *59*, 106–112.

29. Data sheet HSMS-285x. Avago Technologies. Available online: http://www.avagotech.com (accessed on 28 September 2012).

30. Data sheet HSMS-286x. Avago Technologies. Available online: http://www.avagotech.com (accessed on 28 September 2012).

31. Hsu, T.S. Low-frequency excess noise in metal—Silicon Schottky barrier diodes. *IEEE Trans. Electron Devices* **1970**, *17*, 496–506.

32. Hastas, N.A.; Dimitriadis, C.A.; Dozsa, L.; Gombia, E.; Amighetti, S.; Frigeri, P. Low frequency noise of GaAs Schottky diodes with embedded InAs quantum layer and self-assembled quantum dots. *J. Appl. Phys.* **2003**, *93*, 3990.

33. Gomila, G.; Reggiani, L.; Rubi, J.M. Shot-noise suppression in Schottky barrier diodes. *J. Appl. Phys.* **2000**, *88*, 3079.

34. Kleinpenning, T.G.M. Low-frequency noise in Schottky barrier diodes. *Solid-State Electron.* **1979**, *22*, 121–128.

35. Lee, T.H. *The Design of CMOS Radio-Frequency Integrated Circuits*; Cambridge University Press: Cambridge, UK, 2004.

36. Advanced Design System. Agilent Technologies: Santa Clara, CA, USA.

37. Bowick, C.; Blyler, J.; Ajluni, C.J. *RF Circuit Design*; Newnes/Elsevier: Amsterdam, The Netherlands, 2008.

38. Joe, J.; Chia, M.; Marath, A.; Ang, C. Zero Bias Schottky Diode Model for Low Power, Moderate Current Rectenna. In *Proceedings of Defence Electronics Technology Seminar (DETS 97)*, Singapore, Singapore, November 1997; pp. 1–5.

39. Visser, H.J.; Vullers, R.J.M.; Veld, B.O.h.; Pop, V. Remote RF Battery Charging. In *Proceedings of Power MEMS 2010*, Leuven, Belgium, December 2010; pp. 37–40.

40. Nimo, A.; Grgic, D.; Reindl, L.M. Impedance optimization of wireless electromagnetic energy harvester for maximum output efficiency at μW input power. *Proc. SPIE* **2012**, *8341*, 83410W:1–83410W:14.

41. Nimo, A.; Grgic, D.; Reindl, L.M. Electrically small Planner Antenna for Compact Electromagnetic (EM) Wireless Energy Harvester. In *Proceedings of Power MEMS 2011*, Seoul, Korea, November 2011; pp. 310–313.

42. RT-duroid-5870–5880-Data-Sheet; Rogers Corporation: Rogers, CT, USA, 2011.

43. *HFSS*; Ansys Corporation. Ansys, Inc: Canonsburg, PA, USA, 2005.
44. Friis, H.T. A Note on a Simple Transmission Formula. *Proc. IRE* **1946**, *34*, 254–256.
45. Kvaksrud, T.I. *Range Measurements in an Open Field Environment*, *Design Note DN018*; Texas Instrum. Incorporated: Dallas, TX, USA, 2008.
46. *LabView*; National Instruments: Austin, TX, USA, 2009.

Appendix A: Measuring Setup for RF Rectifier Efficiency and Voltage Sensitivity

The measuring setup is as shown in Figure A1.
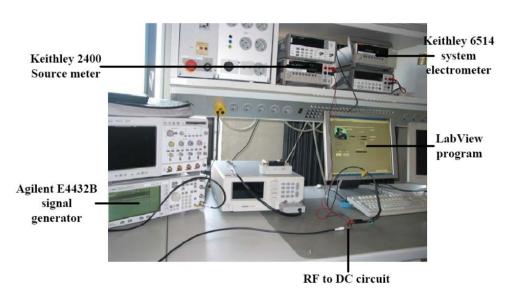
Figure A1. RF to DC Power converter characterization setup.

The RF to DC circuit efficiency and voltage sensitivity measurements were made with a Keithley 2400 source meter and Keithley 6514 system electrometer with an Agilent E4432B signal generator providing 50 Ω RF signal into the circuit board.

The closed circuit current drawn by the RF to DC power converter (*without load*) from the generator is first determined by the Keithley 2400 source meter; then starting from this current, the value of the current is decreased at set intervals to creates virtual load resistances to the circuit for up to a lowest current of 0.1 μA. The 6514 system electrometer is used to measure the output voltage. The number of data point is set through LabView [46] as well as the measurements. Additionally open circuit voltage or at specific loads and frequency sweep can be made through the LabView program. At −40 dBm input power and below, the detected voltages and currents were difficult to measure accurately with the measuring setup; hence measurements were made up to a minimum of -35 dBm input power. The circuit layout for the efficiency and voltage sensitivity measurements is schematically shown in Figure A2.

Figure A2. RF to DC power converter characterization circuit.

STANDARD ANTENNA TERMS AND RELATED FORMULAS

The "IEEE Standard Definitions of Terms for Antennas" (IEEE STD-145) represents a consistent and comprehensive vocabulary suited for the effective communication and understanding of antenna theory. General use of these definitions of terms would eliminate much of the wide-spread inconsistency concerning antenna characteristics, particularly with regard to the basic parameters of gain, beamwidth, polarization and efficiency. For convenience, IEEE antenna terms of general interest are listed here. Wherever these terms appear in this catalog, the definitions given below apply. Other commonly used terms, not covered by the IEEE standard, are shown with an "*."

ANTENNA APERTURE. A surface, near or on an antenna, on which it is convenient to make assumptions regarding the field values for the purpose of computing fields at external points. **Note:** The aperture is often taken as that portion of a plane surface near the antenna, perpendicular to the direction of maximum radiation, through which the major part of the radiation passes.

ANTENNA EFFICIENCY OF APERTURE - TYPE ANTENNA. For an antenna with a specified planar aperture, the ratio of the maximum effective area of the antenna to the aperture area.

* ANTENNA FACTOR. That quantity by which the voltage developed across the output of an antenna is related to the incident field strength in which the antenna is immersed. **Note:** Applicable to low frequency antennas and usually refers to a 50 ohm output.

$$\text{AFE (dB m}^{-1}) = E \text{ (dB V / m) - V (dB V)}$$

AFE = Electric Field Antenna Factor
E = Electric Field Strength at antenna
V = Voltage at terminals of antenna

$$\text{AFH (dB AV}^{-1}\text{m}^{-1}) = H \text{ (dB A / m)} - V\text{(db V)}$$

AFH = Magnetic Field Antenna Factor
H = Magnetic Field Strength at antenna
V = Voltage at terminals of antenna

$$\text{AFE (dB m}^{-1}) = \text{AFH (dB AV}^{-1}\text{m}^{-1}) + 51.53$$

for a plane wave in free space.

$$\text{AFB (dB}\frac{pT}{\mu V}) = B \text{ (dBpT) - V}_o \text{ (dB}\mu V)$$

AFB = Magnetic flux Antenna Factor
B = Magnetic flux at the antenna
pT: picoTesla units
V = Voltage at the terminals of the antenna

$$\text{AFB (dB}\frac{pT}{\mu V}) = \text{AFH (dB}\frac{A}{Vm}) + 2$$

APERTURE ILLUMINATION. The field over the aperture as described by amplitude, phase, and polarization distributions.

APERTURE ILLUMINATION EFFICIENCY. For a planar antenna aperture, the ratio of its directivity to the directivity obtained when the aperture illumination is uniform.
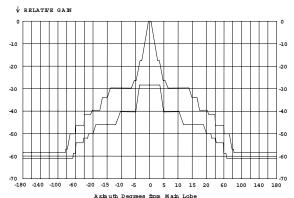
BEAM. The major lobe of the radiation pattern.

CIRCULAR POLARIZATION. It may be either right hand circular polarization (RHCP) or left hand circular polarization (LHCP). The sense of polarization is determined by observation of the direction of rotation of the electric field vector from a point behind the source, RHCP and LHCP correspond to clockwise and counter-clockwise respectively. **Note:** RHCP transmit requires a like polarization to receive.

CO-POLARIZED. The polarization which the antenna is intended to radiate or receive. Also "like polarization".

* CROSS POLARIZATION DISCRIMINATION (XPD). Cross polarization discrimination is the measure of the antenna's ability to differentiate between the vertical and the horizontal polarization of an antenna. This difference, shown in relative signal level, is indicated on directional pattern envelopes (DPE's).

* DIRECTIONAL PATTERN ENVELOPES (DPE'S). In accordance with standard practice, radiation characteristics in any given plane of polarization are measured and plotted using 360-degree polar coordinate systems. The resultant Directional Pattern Envelope is the smoothed composite of all these measurements. The purpose of these DPE's is to emphasize the worst composite condition.

DIRECTIVE GAIN.   In a given direction, 4 times the ratio of the radiation intensity in that direction to the total power radiated by the antenna.

DIRECTIVITY.   The value of the directive gain in the direction of its maximum value.

EFFECTIVE AREA OF AN ANTENNA.   In a given direction, the ratio of power available at the terminals of a receiving antenna to the power per unit area of a plane wave incident on the antenna from that direction, polarized coincident with the polarization that the antenna would radiate.

FAR FIELD REGION.   That region of the field of an antenna where the angular field distribution is essentially independent of the distance from a specified point in the antenna region.

*   FRONT-TO-BACK RATIO.   The ratio of the maximum directivity of an antenna to its directivity in a specified rearward direction.

*   Gain, dBi.   The gain expressed in decibels relative to an isotropic radiator that is linearly polarized.

$$G(dBi) = 10\log(G) \qquad G = 10^{\frac{G(dBi)}{10}}$$

*   GAIN, dBic.   The gain expressed in decibels relative to an isotropic radiator that is circularly polarized.

HALF-POWER BEAMWIDTH.   In plane containing the direction of the maximum of a beam, the angle between the directions in which the radiation intensity is one half the maximum value of the beam.

HALF-WAVE DIPOLE.   A half wavelength antenna, center fed so as to have equal current distribution in both halves. Mounted vertically, it has a doughnut shaped pattern, circular in the horizontal plane. It is an antenna that can be constructed. It has some inherent losses. When used as a gain reference, the half-wave dipole has a power gain of about 1.7 dBi.

* ISOLATION.   Refers to the ability of one port of a dual polarized feed to discriminate against a signal fed into the other port.

ISOTROPIC RADIATOR.   A hypothetical antenna having equal radiation intensity in all directions. Note: An isotropic radiator represents a convenient reference for expressing the directive properties of actual antennas.

NEAR-FIELD REGION.   The spherical region of space between the antenna and the far field region.

NULL.   The region of a radiation pattern, either computed or measured, where the amplitude goes through a minimum value.   Note: (1) It represents the angular position where the phase or the far field pattern crosses the zero axis if the pattern is plotted as a phasor instead of a scalar value.   Note: (2) The region outside the main beam of a directive antenna pattern consists of a series of minor lobes separated by nulls.

PARALLEL POLARIZATION.   The condition where the electric vector is parallel to the local conducting surface. Note: Over the earth, this is usually referred to as being horizontal polarization.

PHASE CENTER.   The location of a point associated with an antenna such that, if it is taken as the center of a sphere whose radius extends into the far-field, the phase of a given component over the surface of that radiation sphere is essentially constant, at least over the portion of the sphere where the radiation is significant.

POLARIZATION. The polarization of an antenna is defined as the polarization of the electromagnetic wave as described by the shape and orientation of an ellipse, which is the locus of the extremity of the field vector, and the sense in which the ellipse is traversed with time. The elliptical locus is called the polarization ellipse and the wave is said to elliptically polarized. Circular polarization and linear polarization are degenerate cases of elliptical polarization.

POWER DENSITY AT A POINT

$$S_{av} = \frac{GP_t}{4\pi r^2}$$

S_{av} = Time average power density in W/m^2
P_t = Power transmitted in watts
G = Antenna gain relative to an isotrope
r = Distance from antenna to point in meters

POWER DENSITY TO VOLTS/METER IN FREE SPACE

E^2 (V / m) = 377 S_{av} (W / m^2)

S_{av} (W / m^2) = E^2 (V / m) / 377

1 V / m = 2.65 mW / m^2

POWER GAIN. In a given direction, 4 times the ratio of the radiation intensity in that direction to the net power accepted by the antenna from the connected transmitter. **Note:** (1) When the direction is not stated, the power gain is usually taken to be the power gain in the direction of its maximum value. (2) Power gain does not include reflection losses arising from mismatch of impedance.

POWER GAIN IN PHYSICAL MEDIA. In a given direction and at a given point in the far field the ratio of the power flux per unit area from an antenna to the power flux per unit area from an isotropic radiator at a specified location with the same power input as the subject antenna.

Note: The isotropic radiator must be within the smallest sphere containing the antenna. Suggested locations are antenna terminals and points of symmetry, if such exist.

POWER GAIN REFERRED TO A SPECIFIED POLARIZATION. The power gain of an antenna, reduced by the ratio of that portion of the radiation intensity corresponding to the specified polarization to the radiation intensity.

POWER TRANSMISSION FORMULAS

$$P_r = P_t \frac{G_t G_r \lambda^2}{(4\pi r)^2}$$

P_r (dB W) = P_t (dB W) + G_t (dBi) + G_r (dBi) - 20 log r - 20 log f + 27.56

P_r (dB W) = P_t (dB W) - AFE$_t$ (dB m^{-1}) - AFE (dB m^{-1}) - 20 log r + 20 log f - 32

P_r = Power received
P_t = Power transmitted
G_r = Gain of receiving antenna
G_t = Gain of transmitting antenna
f = Frequency in MHz, λ = Wavelength
r = distance between antennas in meters
AFE_r = AFE of receiving antenna
AFE_t = AFE of transmitting antenna

RADIATOR. Any antenna or radiation element that is a discrete physical and functional entity.

RADIATION, ELECTROMAGNETIC. The emission of energy in the form of electromagnetic waves.

RADIATION INTENSITY. In a given direction, the power radiated from an antenna per unit solid angle.

RADIATION LOBE. A portion of the radiation pattern bounded by regions of relatively weak radiation intensity

RADIATION PATTERN (ANTENNA PATTERN). A graphical representation of radiation properties of the antenna as a function of space coordinates. **Note:** (1) In the usual case the radiation pattern is

determined in the far-field region and is represented as a function of directional coordinates. (2) Radiation properties include power flux density, field strength, phase, and polarization.

* RADIATION RESISTANCE OF AN ELECTRICALLY SMALL LOOP ANTENNA. The resistive component of an antenna's input impedance that results from the coupling of the antenna to its environment. This resistance dissipates the power that is actually radiated from the antenna.

$$R_r = 20\,(2\pi/\lambda)^4\,n^2\,A^2\ \text{ohms}$$

n = number of turns
A = area of the loop

REALIZED GAIN. The power gain of an antenna in its environment, reduced by the losses due to the mismatch of the antenna input impedance to a specified impedance.

* REALIZED RADIATOR EFFICIENCY. The efficiency of an antenna in its environment reduced by all losses suffered by it, including: ohmic losses, mismatch losses, feedline transmission losses, and radome losses.

RELATIVE POWER GAIN. The ratio of the power gain in a given direction to the power gain of a reference antenna in its reference direction. **Note:** Common reference antennas are half-wave dipoles, electric dipoles, magnetic dipoles, monopoles, and calibrated horn antennas.

RETURN LOSS. The reflection coefficient of a mismatch expressed in decibels. **Note:** Modern swept VSWR techniques actually sense the reflected component which is normalized to the forward component to yield return loss. A 2:1 VSWR is equivalent to 9.5 dB return loss.

VSWR. The voltage standing wave ratio of a component such as an antenna. It is referred to the characteristic impedance of the transmission line being used. **Note:** The most common characteristic impedance is 50 ohms, but 75 and 300 ohms are frequently used in coaxial or twin lines for VHF, UHF applications.

NOTES:

THE HANDYMAN'S GUIDE TO OSCILLOSCOPES (Part 2 of 2)

by Paul Harden, NA5N

Making some advanced measurements with your Oscilloscope

Print as .pdf file
4 pages
8½ x 11 or A4

In Part 1, oscilloscope operation was covered for making basic voltage, time and frequency measurements. In this part, we'll continue with some more advanced uses of a scope, and in particular, how to use a scope for testing and troubleshooting ham radio QRP transceivers in the homebrewer's workshop.

Receiver Filter Bandwidth.

This procedure uses a scope (a DVM can be used with less accuracy) for determining the overall filter bandwidth (or selectivity) of a receiver. It is basically measured by plotting output voltage vs. audio frequency to construct a picture of the filter response.

Connect scope to the receiver audio output (speaker or phone jack); measurement will be based on peak-to-peak voltages (Vpp) on a scope, or rms voltage (Vrms) on a DVM.

Using a signal generator, set the frequency for the band of interest on your radio. For example, on a general coverage shortwave receiver, you might set it for 10 MHz (top end of the 31M band), or perhaps to 7.040 MHz on a 40M ham radio receiver/transceiver. Tune the receiver to the signal generator signal. If you don't have a signal generator, you can also tune to a steady carrier or station to produce a hetrodyne audio "pitch." Tune in the signal to the pitch that causes the maximum peak-to-peak display. Adjust the scope and volume control to produce a 2Vpp display (4 divisions). This is the peak response of the overall filtering stages as shown in **Fig. 13**.

Now determine the audio frequency at this peak response by measuring the time period between cycles and covert to frequency. In the example to the right, the period of one cycle is 1.7mS, which is an audio tone of 750Hz (1/.0017sec). A frequency counter on the output can also be used.
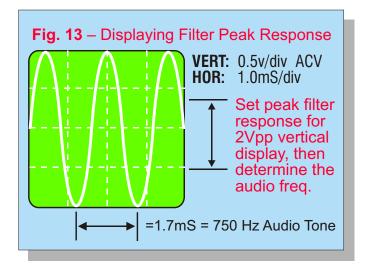
Next, tune the receiver such that the sidetone pitch goes UP in frequency and the peak-to-peak signal will decrease in magnitude. Tune to the point where the signal is exactly 1Vpp on the scope. See **Fig. 12**.
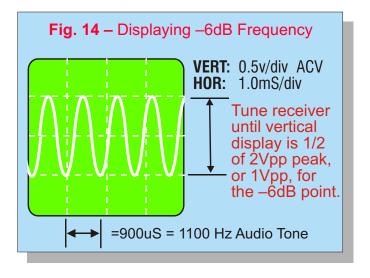
This is the –6dB point of the high end of the filter (20log 1v/2v =–6dB). Determine the frequency of the audio pitch as before. In the example, this is 1100 Hz. Record the data.

From these two data points, the –6dB bandwidth can be estimated. The bandwidth from the filter peak (750 Hz) to the –6dB point (1100 Hz) is 350 Hz. The bandwidth (BW) between the two –6dB points is usually twice this value, or 700 Hz. A filter with a –6dB BW of 700Hz is a mediocre filter for CW reception, and way too narrow for SSB or AM.

Of course you can determine the exact –6dB BW by tuning the receiver back to the 2Vpp peak response, and continue tuning DOWNward in frequency until the audio is again exactly 1Vpp. Determine this frequency and record. In this example, it should occur around 400 Hz if the filter shape is symmetrical.

Plot these three data points on a sheet of graph paper as shown in **Fig. 15** to construct the filter shape. Return to the upper –6dB point (1100Hz in the example) and continue tuning upwards in audio pitch, recording the frequency at 0.5v (–12 dB), 0.25v (–18 dB), 125mV (–24 dB), etc. Everytime you "halve the voltage," it is a 6 dB change. The more points you collect, the more accurate your filter response plot will be.

Fig. 13 – Displaying Filter Peak Response

VERT: 0.5v/div ACV
HOR: 1.0mS/div

Set peak filter response for 2Vpp vertical display, then determine the audio freq.

=1.7mS = 750 Hz Audio Tone

Fig. 14 – Displaying –6dB Frequency

VERT: 0.5v/div ACV
HOR: 1.0mS/div

Tune receiver until vertical display is 1/2 of 2Vpp peak, or 1Vpp, for the –6dB point.

=900uS = 1100 Hz Audio Tone
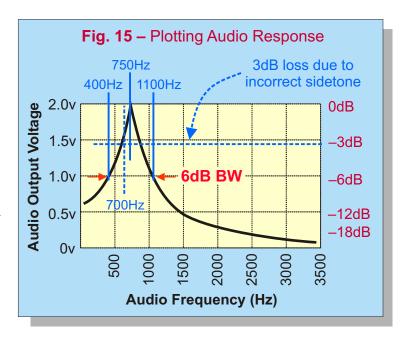
Of interest to proper rig alignment is to repeat the above using the output of the product detector. To maximize the effectiveness of the filtering, the receive offset frequency set by the BFO should be adjusted to the same frequency as the peak frequency response of the audio. In this example, with the peak audio response occuring at 750Hz, if your BFO is set for a sidetone frequency of 700 Hz, you are loosing 2–3 dB, since this is in your filter skirt. This is shown on the response plot in **Fig. 15** by the dashed lines. By adjusting your BFO for a sidetone frequency of 750 Hz, you will pick up 3-4dB of overall gain in your receiver, plus increase the selectivity a bit as well. Why? Because nearby stations, such as one at 800–900Hz tone, could actually be louder than the 700Hz tone signal you are trying to copy, since the gain of the receiver is greater at those tones than at 700Hz, as shown in the plot.
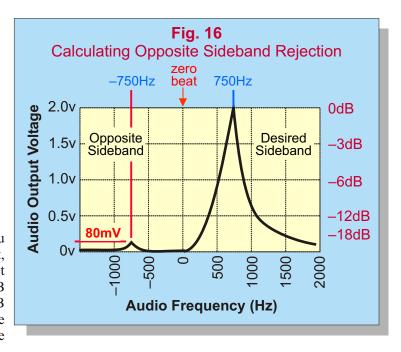
Opposite Sideband Rejection

A superhet receiver is supposed to pass just one sideband and reject the other. Poor opposite sideband rejection could indicate the crystals in your IF filter are not well matched or other problems. It is measured almost identical to plotting the filter response just described. First, you tune the receiver to the test signal to find the peak response frequency, or 750Hz in this example. Set the scope display for 2Vpp. Now tune downward in audio pitch, passing zero-beat, and continue tuning and you should hear the test signal, much weaker, now rising again in tone. This is the *opposite sideband*. Measure the peak-to-peak voltage, if you can. For example, say it is 80mVpp, as shown in **Fig. 16**. Calculate the opposite sideband rejection by:

$$\text{rejection} = 20\log \frac{80mV}{2.0v} = \frac{80mV}{2000mV} = -28dB$$

If you can't hear the opposite sideband, then obviously you have excellent filter rejection. If you can just barely hear it, you may have to increase the sensitivity of your scope (set vertical gain to 20 or 50 mV/div). In this example, –28dB rejection of the opposite sideband is quite good. A –30dB rejection means the opposite sideband is only 1/1000th of the desired sideband, a very suitable attenuation of the opposite sideband.

Fig. 15 – Plotting Audio Response

Fig. 16
Calculating Opposite Sideband Rejection

These tests are important to perform on your rig for documenting it's current performance, and repeated periodically to detect unfavorable changes or for troubleshooting when a problem is evident. For homebrewing, these tests can allow you to evaluate different circuits or when experimenting with different components or part values.

Oscillator Phase Noise
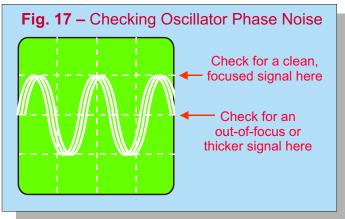
When homebrewing a basic oscillator circuit, such as a VFO, looking at the oscillator output on a scope can reveal several problems. One is to check for excessive phase noise. Phase noise is small variations of the oscillator frequency that causes power in the close-in sidebands, usually measured on laboratory equipment within 100KHz, or even within 10 KHz.

To check for oscillator phase noise, connect the oscilloscope to the oscillator output, loading the output of the oscillator as little possible. Most scopes have sufficiently high input impedances where this shouldn't be a problem, but some cheaper scopes can load an oscillator circuit. If you suspect your scope is loading the oscillator, couple the scope to the circuit with a small value capacitor, less than 20pF.

Display 2–3 cycles of the oscillator output as shown on the scope display shown in **Fig. 17**. Properly focus the scope and carefully observe if the waveform appears in focus at the peaks, but slightly out-of-focus at the zero-crossing points, that is, on the rising and falling edges of the sine wave.

If it appears out-of-focus, this is excessive phase noise jittering the signal and "smearing" the waveform along the time (horizontal) axis. Extreme phase noise may show 2-3 sine waves very close to each other, as shown in the exaggerated waveform to the right — assuming you have your scope properly triggered.

Phase noise is random, instantaneous changes in the oscillator

Fig. 17 – Checking Oscillator Phase Noise

Check for a clean, focused signal here

Check for an out-of-focus or thicker signal here

frequency that smears the display. If you can see this on a scope, the phase noise is excessive! If you can't see it, it doesn't mean the oscillator has no phase noise (all oscillators have some phase noise), it just means it is not excessive enough to see on a scope. A scope is not a good instrument for checking phase noise, but for homebrew circuits, it is a check to ensure you do not have a serious oscillator problem.
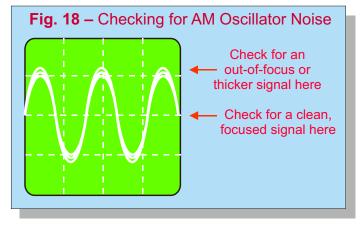
Excessive oscillator phase noise in receivers can cause IMD products and noise in the audio range at the output of the mixer(s), including the product detector. In a transmitter, excessive oscillator phase noise will put power in the close-in sidebands of your carrier, not only wasted power for lower transmitter efficiency, but may produce strange sounds (buzzing or chirping) to the receiving station. A few causes of phase noise are excessive current in the oscillator transistor, low-Q coil(s), high dissipation in the tuning caps or poor power supply filtering at the oscillator frequency.

AM (Amplitude) Noise
Another oscillator problem may be AM noise, or amplitude modulated noise. It is an opposite effect on a scope when displaying the oscillator output — the sine wave appears out-of-focus or thicker at the peaks, and in-focus elsewhere as shown in **Fig. 18**.

If you detect AM noise, slow down the scope's sweep rate to the audio frequencies or slower to see if you can notice a lower frequency component. A common cause of AM oscillator noise is 60 Hz from the power mains leaking into the circuits. This is particularly true if using a power supply off of 120v 60 Hz, or sometimes it can be due to the AC lighting above your head! If the AM noise seems to be at the same frequency as the audio output tone, it means audio is getting into the Vcc bus, likely due to poor bypass filtering at the audio amplifier stages (particular if using an LM386 or similar).

If you can't find a low frequency component, the AM noise may be random, which may indicate poor voltage regulation, a noisy voltage regulator, or perhaps a circuit in VHF oscillation. If the

Fig. 18 – Checking for AM Oscillator Noise

Check for an out-of-focus or thicker signal here

Check for a clean, focused signal here

AM noise seems to occur on key-down only on a transmitter, the transmit current may be loading the power supply, the voltage regulator is under-rated, or just simply loading the oscillator. In the case of loading the power supply or regulator(s), perhaps a separate voltage regulator or zener circuit should be used, dedicated for the oscillator(s). In the case of transmit loading, adding a buffer amplifier or emitter follower to isolate the load from the oscillator may help.

Much of this can be diagnosed also with the scope, by looking at the AC ripple on the DC power lines. You should have less than 50mV of any AC component on the 8-12v DC wiring, whether 60Hz, audio or RF. If >50mV, then additional low or high frequency filtering on the DC power is needed.

Monitoring Transmitter RF Output
RF Power (in watts) is E^2/R, where E is in rms and R=50 . The voltage displayed on the scope (peak-to-peak) must be converted to rms by Erms=.707(Epp/2). To measure properly, the transmitter should be on a 50 dummy load using the normal hi-Z scope input. If your scope has a selectable "50 "input, it can be the transmitter dummy load directly, providing it can tolerate the 50Vpp input. (Always know what the maximum p-p input voltage your oscope will tolerate. It is often stated on your scope at the vertical channel inputs. 50Vpp to 100Vpp are typical).
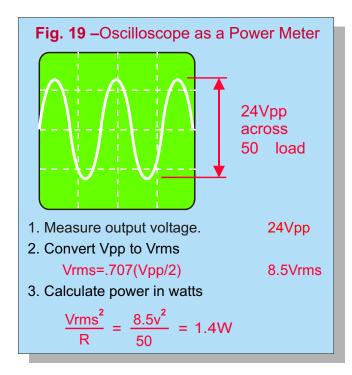
Measuring Transmitter RF Output Power

Figure 19 shows how you can use an oscilloscope to fairly accurately measure output power from a low power transmitter (generally 10W or less with a 1X probe and 20-30W with a 10X probe). Connect transmitter output to a 50 dummy load. Connect the scope lead to the transmitter output (after the low pass filters) or directly to the dummy load. Do not connect the scope leads to the collector or drain of the final PA transistor. The displayed voltage in this case will be erroneous.

In the example, the output transmit voltage across the dummy load measures 24Vpp. Convert this voltage to Vrms, then use the equation (Vrms squared divided by the load resistance) to calculate the power in watts. In the example, 24Vpp, the output power is 1.4 watts. A 5W QRP transmitter should produce about 45Vpp.

The accuracy of your power measurements depend upon the condition of your oscilloscope:
1) Ensure your scope's vertical sensitivity (volts/division) is properly calibrated (see Part I)
2) Ensure your vertical sensitivity is properly calibrated with the 1X or 10X probe you are using.

3) Ensure that the transmitter frequency is well within the bandwidth of your scope. A 100 MHz scope should give reliable readings on all HF bands. A 50 MHz scope should be reliable on all HF bands except slight errors at 28MHz.

Fig. 19 –Oscilloscope as a Power Meter

24Vpp across 50 load

1. Measure output voltage.   24Vpp
2. Convert Vpp to Vrms
   Vrms=.707(Vpp/2)   8.5Vrms
3. Calculate power in watts

$$\frac{Vrms^2}{R} = \frac{8.5v^2}{50} = 1.4W$$
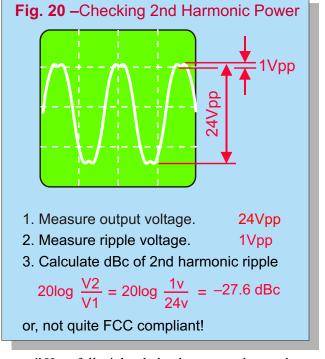
Checking Transmitter Purity

Phase Noise of the transmitter is measured identically to the phase noise checks on page 2–3. With the scope properly triggered and focused, you are looking for an out-of-focus "fuzziness" around zero crossing. If it appears phase noise exists, the fault is generally at the beginning of the transmit chain. That is, seldom does the PA transistor add phase noise; it's merely amplifying what it is given. Phase noise on the transmitter output is more likely due to the transmit oscillator, or the transmit mixer if used.

Harmonic Power can also be detected on a scope while looking at the transmitter RF output power. In **Fig. 20**, notice the "dips" or the two "peaks" on the top and bottom of each sine wave. This is caused by excessive 2nd harmonic output power. The rule-of-thumb is – if you can see any 2nd harmonic power (a dip or flattening at the peaks), then you are right at or exceeding the –30dBc FCC harmonic specification. A clean sine wave implies FCC compliance, providing the 2nd harmonic is within your scope's bandwidth. Obviously, a spectrum analyzer should be used for accurate harmonic evaluation, but a scope can be used to indicate if you have a problem. If you build your own transmitter, it is your responsibility to ensure it is compliant. An approximate method of estimating 2nd harmonic attenuation is shown.

–30dBc means the 2nd harmonic power is 1/1000th of the fundamental power. For a 5W transmitter, this means the 2nd harmonic power should be 5mW or less, or about 1Vpp. In displaying a 5W signal on a scope (45Vpp), it would be difficult to resolve much less than a 1Vpp dip or ripple.

Fig. 20 –Checking 2nd Harmonic Power

1Vpp

24Vpp

1. Measure output voltage.   24Vpp
2. Measure ripple voltage.   1Vpp
3. Calculate dBc of 2nd harmonic ripple

$$20\log \frac{V2}{V1} = 20\log \frac{1v}{24v} = -27.6 \text{ dBc}$$

or, not quite FCC compliant!

Conclusion. This concludes the NA5N "Handiman's Guide to Oscilloscopes." Hopefully, it has helped you to understand your oscilloscope to make simple measurements, and with some practice, perform the more advanced measurements discussed. An oscilloscope is a very powerful tool that is invaluable on the homebrewer's workbench. If there is a particular measurement you feel I omitted, please let me know and I will add it when time permits.

73, Paul Harden, NA5N (na5n@zianet.com or pharden@nrao.edu)

THE HANDYMAN'S GUIDE TO OSCILLOSCOPES (Part 1 of 2)

by Paul Harden, NA5N

Getting Acquainted with your Scope and making some measurements

Updated May 2004 with actual oscope waveforms
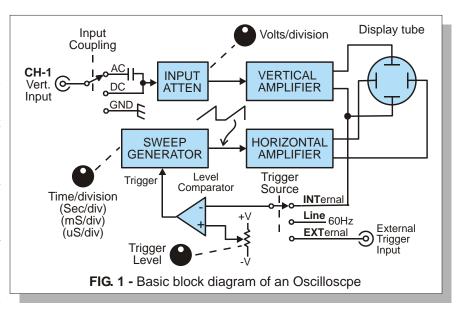
When in the course of human events, it becomes necessary to look at neat signals floating around your radio, you go to a hamfest and buy that $50 o-scope. Now what? This two part article will attempt to explain basically how an oscilloscope works, operator functions, basic measurements, and some advanced applications. An o-scope is a powerful tool in any shack – even a real "cheapie" with limited bandwidth.

HOW AN OSCILLOSCOPE WORKS

A block diagram of a typical o-scope is shown in **Fig. 1**. The test probe usually plugs into the scope via a BNC connector, then passes through a switch to determine whether the input signal will be dc or ac coupled (to remove any dc component). Often this switch will have a "ground" position for setting the zero-volts reference. Next is the input attenuators. The vertical input amplifier is quite sensitive, designed for 20-50mV of input. For larger input voltages, the signal is applied to attenuators comprised of simple voltage dividers. This is the first area of concern for cheap o-scopes, as the input attenuators may not be very linear or accurate. For example, if you appy a 10Vpp signal on the 10v/division setting, the signal should be 1 division high. Switching to 1v/div., the signal should be 10 divisions (usually full-scale) high. If it is not exactly 10 divisions, the attenuator for that setting needs adjusting. Some scopes have internal adjustments for fine-tuning each attenuator setting.

Following the attenuators, the signal is applied to the first vertical amplifier, which converts the input to a differential signal. This differential signal is amplified up to high voltages for the oscilloscope deflecting plates – moving the beam up and down (in the vertical axis).

The sweep generator is usually a constant current source charging a capacitor to make a sawtooth waveform that eventually deflects the beam in the horizontal axis. The frequency of the sawtooth determines how fast the beam travels from the left to the right side of the tube, and is controlled by the sweep control, usually calibrated in seconds, milli-seconds or micro-seconds per division. This is the second area of concern for an oscilloscope – how linear the sawtooth waveform is generated. For example, a sawtooth with a nonlinear ramp will cause the signal displayed in the central portion of the tube to be expanded or compressed compared to the signal at the ends.

FIG. 1 - Basic block diagram of an Oscilloscpe

The sawtooth ramp is amplified to high voltages, applied to the oscope tube, to deflect the beam from left to right. An important task of an oscope is when the horizontal deflection begins. Normally a switch labeled "Trigger Source" determines what initiates the sawtooth ramp. In the "Internal" position, a sample of the input signal (in the vertical amplifiers) is sampled, with a variable resistor setting the level. When the input signal exceeds the "Trigger Level," a pulse is generated to start the sawtooth ramp and hence the horizontal sweep. The purpose of triggering is to keep the input waveform synchronized to the sweep so it appears stationary on each sweep. The trigger source usually has a "Line" position, which simply triggers the sweep off of 60Hz from the power supply. This synchronizes the sweep to the AC power frequency and is useful for checking television signals, which are synchronized to the power mains. Also, an "External" position may be present, which connects an external input signal (via a BNC connector) to trigger the sweep generator.

Other features your oscope may have are two vertical channels for dual trace operation, various modes to display both waveforms (alternate, chopped, A+B added, etc.), delayed sweep features, dual sweep time bases, built in calibrators, etc.

CALIBRATING YOUR OSCILLOSCOPE
The first thing you should do upon acquiring an o-scope is to check its calibration.

The vertical amplfiers can be checked with a known voltage source or 9v transistor radio battery. Measure the output voltage of the battery with an accurate voltmeter. Let's say it just happens to be +9v exactly. Set the input coupling to ground (0v) and move the trace to the bottom division. Switch the input coupling to DC and set the attenuators to 1v/div. The deflection should be 9

divisions. Switching to 10v/div., deflection should be 0.9 divisions. Internal to the oscope (or perhaps accessible from the outside) are adjustments for the vertical amplifier gain. Adjust this for 9 divisions of deflection in the 1v/div. range. Procedure can be repeated with a 1.5v flashlight battery (assuming you know the exact voltage from a DVM).

The horizontal amplifiers should be checked/calibrated using a signal generator. For example, a 1MHz signal has a period of 1uS. Setting the sweep rate to 1.0uS/div., a 1MHz signal should take exactly 1 division per cycle. Set the horizontal width control properly to ensure the beam starts at the first division and ends at the last division. If the sweep rate appears incorrect, an internal adjustment (Sweep gain or similar) can be set for proper display of the test signal.

The main operator controls are:
• **Intensity** - controls the brightness of the beam. **NOTE: Too bright a beam can damage to the CRT tube!**
• **Focus -** adjusts the beam for the thinnest and sharpest display.
• **VERT & HOR Position -** controls the vertical and horizontal position of the display respectively
• **VERT V/div -** controls the vertical sensitivity of the display, i.e., how many volts (or mV) per division.
• **HOR Sweep Speed -** sets the horizontal sensitivity, i.e., how many mS or uS per division.
• **VERT & HOR vernier -** allows the vertical and horizontal sensitivity settings to be varied in small steps.

Other adjustments you may find on your scope are:

Astigmatism - With the scope intensity and focus properly set, this adjustment compensates for the curvature of the CRT tube by making it in-focus across the sweep. If your trace is out-of-focus in certain areas, but in-focus elsewhere, the astigmatism needs to be adjusted. See *Fig. 2*.

Trace Rotation - is a small coil around the CRT that skews the trace to ensure it is perfectly horizontal. On scopes without this adjustment, the trace is leveled by physically rotating the CRT to align the trace to the graticle grid. See *Fig. 2*.

DC BAL (DC Balance) - is a dc offset in the vertical amplifiers that causes a shift in the trace baseline when changing vertical scales. It is most obvious when measuring ac voltages. For example, you are displaying a 10Vpp sine wave, centered on the center graticle, at 2v/div. Changing to 5v/div, the sine wave shifts off the center graticle ... that is, it assumes a dc bias error. The DC BAL is adjusted until the shift no longer occurs when changing vertical scales.

HV ADJ. - is the high voltage that controls the intensity of the trace. Turn up the **Intensity** control to its brightest position, then adjust the HV ADJ for a trace slightly brighter than normal intensity. The **Intensity** control now has the proper range. The HV ADJ might have to be re-adjusted to acquire proper focus.

NOTE: Very bright trace displays can cause permanent damage to the CRT, particularly on a well-used scope.
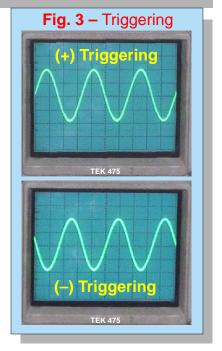
Fig. 2 – Effects of Astigmatism & Trace Rotation

Effects of **ASTIGMATISM** (Inconsistent focus)

Effects of **TRACE ROTATION** (Trace not level)

LET'S MAKE SOME MEASUREMENTS

It is assumed you have your scope relatively calibrated and familiar with the front panel controls. The sample o-scope displays are based on eight vertical and ten horizontal divisions on the CRT screen, typical to most oscilloscopes. Most waveforms are actual displays of the signal cited, photographed from my trusty Tektronix 475 oscilloscope.

First ... a word on TRIGGERING.

Most oscilloscopes have a knob or two for "Triggering." This tells the oscilloscope when to start the sweep. When the **Triggering Slope** is placed in the **(+)** position, the scope will begin its trace when the *input* signal goes positive. Likewise, when **(–)** triggering is selected, the trace will begin when the input signal goes negative, as shown in **Fig. 3**. Often there will be the option to chose the **Triggering Source**, such as "CH.1" or "LINE." Line means the scope is triggered off the 60Hz line voltage, and is useful when synchronizing on television signals or looking at 60Hz power supply noise. CH.1 or CH.2 means the scope will trigger off the signal on channel 1 or 2 respectively. **Trigger Level** is at what voltage of the input signal triggering begins. For example, if set high, triggering may not begin until the input signal reaches several volts. When set around zero, it will trigger the moment the signal goes positive (if set for (+) triggering). This setting can be troublesome if noise exists on the signal. Adjust for stable triggering.

Fig. 3 – Triggering

(+) Triggering

TEK 475

(–) Triggering

TEK 475

DC Voltages.

Say you want to check the transmit-receiver (T-R) switch in your QRP rig, or other digital signal. See *Fig. 4.* The key line is the input to the HCT240 inverter to form the 0v TX– on key- down and the 0v RX– on key- up. This switches the rig between transmit and receive (T-R Switch). It is a logic function, that is, a voltage to represent ON or OFF.

Place the scope lead on pin 13 at 10v/div. and you should see the waveform like the top trace in *Fig. 4* ... about +6v on key-up and 0v on key down. Move the scope lead to pin 7 and you should see 0v on key-up and about +8v on key-down (bottom trace). If the output does not go "HI" (+8v) on key-down, or does not go to a solid "LO" (<1v) on key-up, the inverter is not working properly. (It's busticated). Many shortwave receivers use similar schemes for switching filters or attenuators.

While this test could be done with a DVM, the integration time is slow, requiring long key-downs to get the voltages. A scope will also show you how clean the switching is, or if there is an ac voltage (or RF noise) riding on the T-R voltage.

Scopes are thus good dc voltmeters, with about a 5% reading accuracy.

AC Voltages.

Here is where an oscilloscope pays for itself by making AC voltage (and frequency) measurements. You must remember, AC voltages are displayed on a scope as *peak-to-peak voltages*, while a voltmeter measures in *rms*. RMS voltages are about 1/3 the p-p voltage read on a scope, or specifically:

$$Vrms = \tfrac{1}{2}(.707 \times Vpp) = 0.354 \times Vpp$$

For example, let's measure the output voltage and frequency from the sidetone oscillator in your QRP rig. Place the scope lead on the audio amplifier output. On key-down, you get the waveform shown in *Fig. 5*. The transmit sidetone audio is 1.9Vpp.

AC Frequency Measurement.
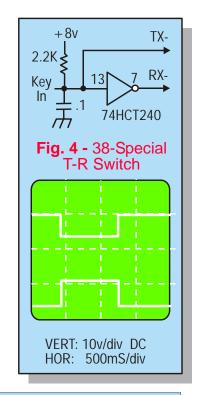
With this waveform, we might as well see what frequency our sidetone or transmit-offset frequency is. Most operators prefer the sidetone to be about 700–750Hz. Trigger the scope for a stable waveform and set the time-base (sweep) to display 2 or 3 cycles, as shown in *Fig. 6*. Center the waveform between two horizontal divisions so zero volts on the waveform is on a graticle line, then move the horizontal position so the first "zero– crossing" is also on a division line.
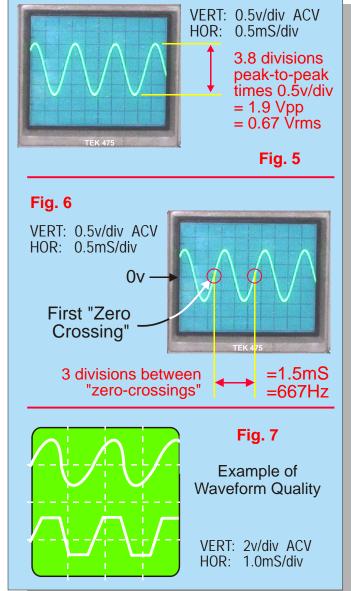
Measure the time it takes to make one complete sine wave from one zero-crossing to the next. In this example, it is 1.5 divisions, at 1mS per division, or 1.5mS. Frequency is simply the reciprocal of time, such that the sidetone frequency is:

$$f = \frac{1}{t} = \frac{1}{1.5mS} = 667 \text{ Hz}$$

For some, this may be about right. For others, this may be a little low to your liking. To raise it to 700Hz, calculate the time period of 700Hz (1/700 = 1.4mS). At 1.0mS/div, you can adjust your sidetone or transmit offset until zero-crossings for a single sinewave is 1.4 divisions. This will be about 700 Hz. (Sidetone may not be adjustable on some rigs).

All frequency measurements are made in this fashion, by measuring the distance between zero-crossings (or from one peak to the next) and converting the time period to frequency. This should emphasize the importance of ensuring your sweep speed is calibrated; as any error in the time base will cause a corresponding error in the accuracy of your time or frequency measurements.

Fig. 4 - 38-Special T-R Switch

VERT: 10v/div DC
HOR: 500mS/div

VERT: 0.5v/div ACV
HOR: 0.5mS/div

3.8 divisions peak-to-peak times 0.5v/div = 1.9 Vpp = 0.67 Vrms

Fig. 5

Fig. 6

VERT: 0.5v/div ACV
HOR: 0.5mS/div

0v

First "Zero Crossing"

3 divisions between "zero-crossings"

=1.5mS
=667Hz

Fig. 7

Example of Waveform Quality

VERT: 2v/div ACV
HOR: 1.0mS/div

Quality of the waveform is another feature of a scope that is unsurpassed since you are "seeing" the waveform in real time. Two examples of waveform quality are shown in *Fig. 7*.

The top trace shows the sidetone frequency with distortion, perhaps due to improper time-constant on the coupling capacitors or improperly biased audio amplifiers. The bottom trace would be a raspy sounding side tone, due to the amplifier being over-driven and in compression (clipping). The o-scope is an invaluable tool for detecting and diagnosing such impurities in the signal quality.

MORE NIFTY MEASUREMENTS

Amplifier Gain.

The gain of an amplifier can be measured in terms of voltage or decibels (dB). For voltage gain, it is simply Vout/Vin of the amplifier. For example, if the input is 1Vpp and the output is 4Vpp, then the amplifier has a voltage gain of 4.

Gain in dB is often more useful and is how the gains of amplifiers are usually expressed. With dB's, every-time you double the AC voltage, you add 6dB of gain. It is the **ratio** of output to the input, and this **ratio** is easy to measure on a scope.

It is often easier to start with the output. Set the vertical amplifier gain to display the amplifier **output** as a full-scale signal as shown in *Fig. 8*. Now move the scope probe to the amplifier **input** without disturbing the scope gain. You will of course have a much smaller signal, and the ratio of the input to the output will be the gain in dB. In our example of using eight divisions for full-scale, then four divisions would be 6db, 2 divisions 12dB, etc. as shown in **Fig. 9.** You may want to add your own dB scale along your scope display to remind you of this relationship. Note: this is *voltage gain* (Av=20log x Vout/Vin). In this example, with 4Vpp output and 1Vpp input (Av=4), then the gain is dB=20log(4) = 20(0.602) = 12dB, or as shown directly on the CRT tube. Since this is a relative measurement, the absolute Vin or Vout voltage does not need to be determined.

Insertion Loss.

In some circuits, such as filters or attenuators, the *loss* in the circuit needs to be measured, and like circuit gain – expressed in dB. The loss through a circuit is called the *insertion loss*. It is determined in the same way as amplifier gain just presented, except start with the input (the highest AC voltage) as the full-scale or reference display, then measure the output AC voltage (the lowest level). **The ratio is the insertion loss in dB.**

For example, with a signal generator connected to your receiver, you want to measure the insertion loss through the IF crystal filter. At the filter input, you can just barely squeek out 2 divisions of input signal on your scope at its most sensitive setting. The output from the crystal filter is 1.5 divisions. The insertion loss would be 20log(1.5/2.0 div.) = –2.5 dB. If the output were only 1.0 division (50% reduction), the insertion loss would be 6dB.
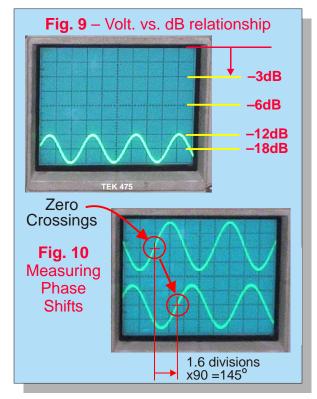
Measuring Phase Shifts.
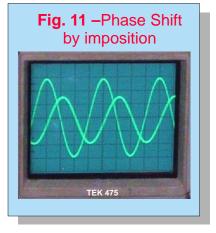
Phase relationships between two signals at the same frequency can be measured with 2-5° accuracy with a scope, although more suited for a dual-trace scope. The reference signal is applied to CH. 1 and the signal to be measured to CH. 2. For proper phase measurements, ensure your dual trace display is in the **chopped** mode, not **alternate** mode for proper phased referenced triggering.

FIG. 8 – Full Scale Signal Display

Actual display
Tektronix 475

Fig. 9 – Volt. vs. dB relationship

–3dB
–6dB
–12dB
–18dB
TEK 475

Zero Crossings

Fig. 10 Measuring Phase Shifts

1.6 divisions
x90 =145°

Fig. 11 –Phase Shift by imposition

TEK 475

There are many methods to do this. One is to stretch out the signal so it takes 4 horizontal divisions, such that each division is 90° of phase, as shown in **Fig. 10**. By measuring from a common point on one signal (zero-crossing or from peak-to-peak) to the next, the phase can be measured. For example, say you are making a phased-array antenna in which one feedline must cause a 90° delay. You calculate the electrical length for a ¼ [L=(246/f) x Velocity Factor] and cut the coax to that length. You are now working on blind faith that you have exactly 90°. With a scope, you can measure it fairly accurately by injecting a signal into one end with a signal generator (at the frequency of interest) and a 50 load on the other. Connect the scope CH.1 to the coax (signal) input and CH.2 to the load end and measure the phase. In the **Fig. 10** example, the CH.2 signal is delayed by 1.6 divisions, at 90° /div is 145. *Your delay line is too long!* Cut off an inch or two at a time until the CH.2 signal is 90° from CH.1 for precise tuning of the delay

line. (While departing from o-scopes for a moment, the sharp null of a phased array is astounding when exactly 90° delay is achieved. More than $10\text{-}15^o$ in error causes a very "mushy" null with little difference over a single vertical antenna. Most errors in achieving exactly 90° by the "measure-and-cut" method are due to uncertainties in the stated velocity factor of the coax).

Another method is to superimpose the two signals on top of eachother. Make one signal larger than the other so you know which one is what, as shown in **Fig. 11**. In this example, the smaller signal lags the larger signal by about 100°, estimated by where they cross. For more accurate determination, use the time base to measure the time period of one cycle (T1), then the time period one signal lags (or leads) the other (T2). The phase shift is then $=(T2/T1)\text{x}360^o$.
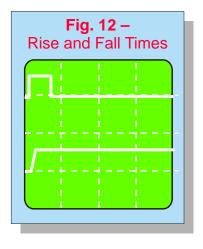
Phase measurements can be made on a single trace scope as well. First, connect the reference signal, uisng a BNC "T," to both the *external trigger* and the normal *vertical input*. Adjust the trigger level so the zero-crossing occurs at the beginning of the trace (left-hand graticle). Remove the reference from the vertical input, but not the e*xternal trigger,* and apply the signal to be tested to the vertical input – without altering the time base or trigger level. The distance of zero-crossing of the test signal is from the left-hand graticle can now be measured to determine the phase, though with slightly less accuracy than using a dual-trace scope.

An interesting experiment is to measure the phase shift of the audio signal at different frequencies as it travels through the stages in a CW, SSB or AM active filter. What is the phase shift of the wanted vs unwanted frequencies?

Measuring Rise and Fall Times.

In digital circuits, it is sometimes important to know the rise and fall times of a signal through a gate. In amateur radio transceivers, this same interest could be applied to how fast the T-R switch switches. On key-down, if the transmitter turns on slightly before the receiver is turned off, it can produce an annoying "thump" in the receiver. Rise and fall times are measured by triggering on the edge of the signal of interest, then increase to a faster sweep speed to measure the time it takes the signal to reach 90% of its final level. The signal to be measured is shown in **Fig. 12** on the top trace, and the expanded version on the bottom. For proper rise times, the signal being measured should be well within the bandwidth of your scope and using a low capacity probe.

For example, in **Fig. 12** (bottom trace), the rise time is about 1/4th of a division. If the sweep speed is 100nS/division, the rise time would then be about 25nS.

Fig. 12 –
Rise and Fall Times

USING LIMITED BANDWIDTH SCOPES

Today's scopes have 200–500MHz bandwidths. Likely your scope is much less than that. A limited bandwidth scope is still very useful to the amateur or homebrewer. Say the bandwidth of your scope is 5MHz. This does not mean you can't see 7MHz signals. It just means the peak-to-peak value has lost meaning, and will likely be very weak, since it is beyond the bandwidth of the scope. (Like other bandwidth measurements in electronics, the "bandwidth" of a scope is usually based on the "3dB bandwidth." That is, at the maximum bandwidth, you are already at the –3dB point, or a 25% reduction in the peak-to-peak voltage display). You can still resolve individual cycles higher than the cited bandwidth to a certain degree and make gain and phase measurements, since they are based on *ratios.*

Most of the examples in this article explore many regions of a communications receiver or ham transceiver without the benefit of any great bandwidth. Experiment with your scope to learn its limitations. *Use a good scope probe and make measurements with a good ground to get the most out of the bandwidth you have.*

For the homebrewer building circuits in the HF bands, a 50 MHz scope **with good calibration** will yield fairly accurate measurements up to 30 MHz with little concern for accuracy. The old 465 or 475 series of Tektronix scopes, with 100/200 MHz bandwidths, make an excellent oscilloscope for the amateur or experimenter. They can often be found at hamfests today for $100–150, and tend to maintain a fairly good calibration almost regardless of how much use they have seen.

In Part 2 - we'll probe (bad pun) into some advanced measurement techniques, even with a simple scope ... such as measuring sideband rejection, filter responses, VCO phase noise, etc. (and what it all means).

 72, Paul NA5N

na5n@zianet.com
pharden@nrao.edu

NIC NAC TIC TAC CRYSTAL RADIO SET - AN IDEAL PRIMARY SCHOOL PROJECT...

Technology (/technology/) > Science | by mk484 (/member/mk484/)   Follow

10,750   ♥ 83   16

Posted Jul. 4, 2014 | ⓒ PUBLICDOMAIN

(https://cdn.instructables.com/FJT/FTE6/HWZ9T0AW/FJTFTE6HWZ9T0AW.LARGE.jpg)
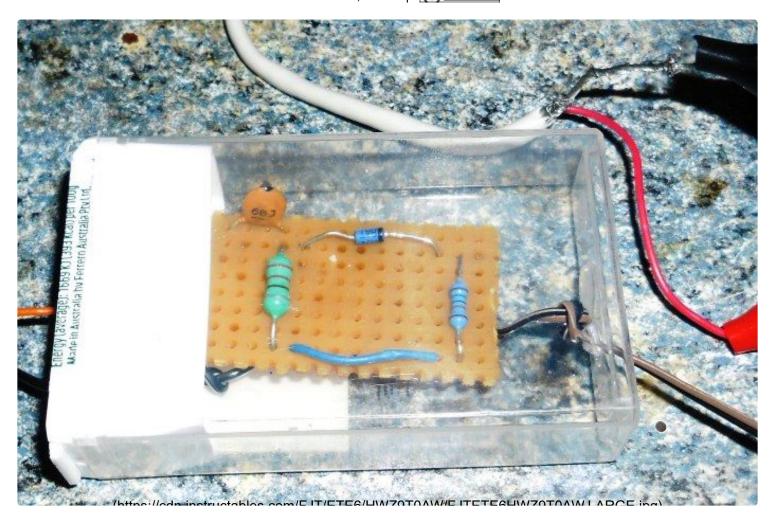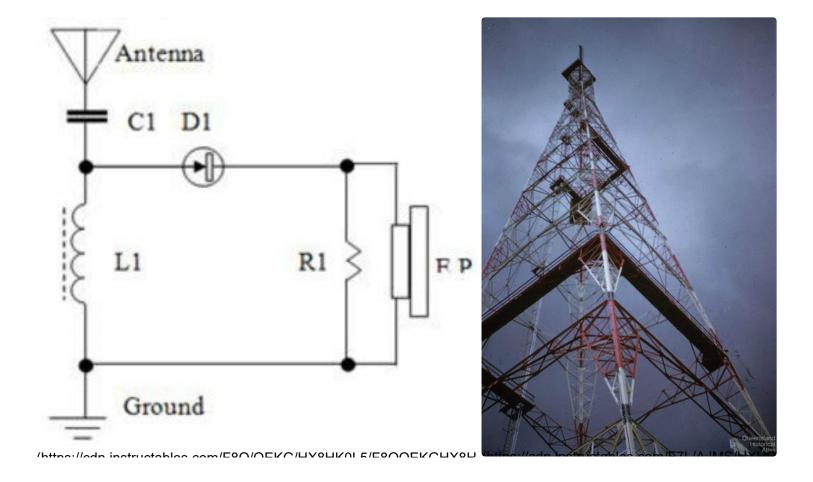
Hi there! This Instructable is all about building a basic crystal radio set that is so simple to build and understand, that a child could do it - with help from mum or dad - or even at school, as a class project. Parts can be bought "off the shelf" at Jaycar Electronics and other suppliers, or purchased online via Ebay and Paypal.

I built my first ever crystal set when I was 9 years old, but that was a different world. We're going to build a 21st century style radio with all modern components which are relatively easy to obtain at your local electronics store or on the Internet.

The radio is contained inside a Tic Tac box, and is called "The Nic Nac Tic Tac Radio". It is built on a square of matrix board, with the component leads poked through the holes in the board, and joined underneath. It is a very simple crystal set, designed to initially receive only one station, but you can expand the tuning range with the addition of only one extra component.
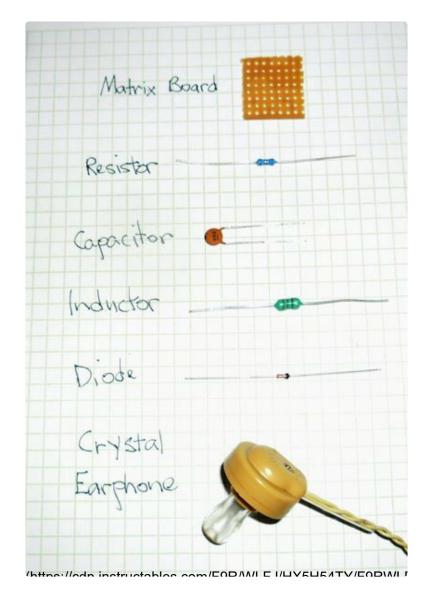
The Tic Tac radio can be constructed by simply twisting the component leads together, underneath the matrix board and attaching antenna/ground wires as well as the earphone wires to the board, using that same technique. You don't have to

solder any wires, but there is another way to join them all together - more on that later.

The Tic Tac Radio is a very easy set to put together - you don't need to wind bulky coils or use heavy tools and screw strange looking parts onto a large breadboard. See step 1 for the parts list you will need.

Add Tip   Ask Question

Step 1: Parts and Tools You Will Need

You will need to acquire the following components from a local electronics shop or an online electronics business.

I am adding the catalogue numbers for Jaycar Electronics, so if you live in Australia or North America, you may be able to visit a Jaycar store and buy the parts over the counter, if not then you may be able to do an online mail order via Paypal:

Parts List

Resistor - 47k - yellow-purple-black-red and brown - RR 0612 (pkt of 8)

Capacitor - 68pF - ceramic x 2 - RC 5322 (pkt of 2) and a 100 or 120 pF value as well for experiments.

Inductor - 220 uH - red-red-brown silver - LF 1538 (resistive type)

Polyvaricon tuning capacitor - 220 pF - RV 5728 - with knob and mounting screws

Diode - BAT46 - ZR 1141 (You can also use a 1N34A Germanium Diode too if you have one at home)

Ceramic Earphone - AS 3305 *

A 25 meter roll of yellow hook up wire for the Antenna wire and a 3 meter length of wire for the Ground wire.

Please note that some Jaycar parts come in multiples of 2 or more per packet. And please note the following:

*A normal crystal radio earphone is OK, but if you can't get one of these, or if the one you bought goes dead (as they sometimes do,) you can use a substitute, such as the Murata PKM44EW passive transducer (see picture above) which is available from an old Telstra TF200 touchphone, (the one on the left in the diagram above,) or an equivalent, such as the ARIO transducer, from an old Telstra T1000 pushbutton phone.

The ARIO unit is soldered to the phone's pc board so you'll need to be able to unsolder the three mounting pins underneath the board, or find someone in the neighbourhood who is able.

Take the back off the TF200 (if you've obtained one of these phones,) and you'll see a black disc shaped object 2" round by 1/2' thick - with a red and black wire. Unplug the wires from the circuit board, and unscrew any retaining screws and remove the transducer. Cut the mini plug off, carefully strip the insulation from the ends of the wires and extend them by about 18" with 2 thin lengths of hookup wire. These pietzo devices make good earphones for crystal sets and can be housed in an old pair of ear muffs.

Miscellaneous Materials:

A Tic Tac box (smaller size)

A piece of matrix board at least 7 holes long by 8 holes across. Cut the board to fit neatly inside the Tic Tac box.

A short length of 2 differently coloured wires 60 mm in length and 2 crocodile clips with red and black plastic covers

A length of antenna wire at least 25 metres long and a 3 metre length of a different colour for the Ground wire.
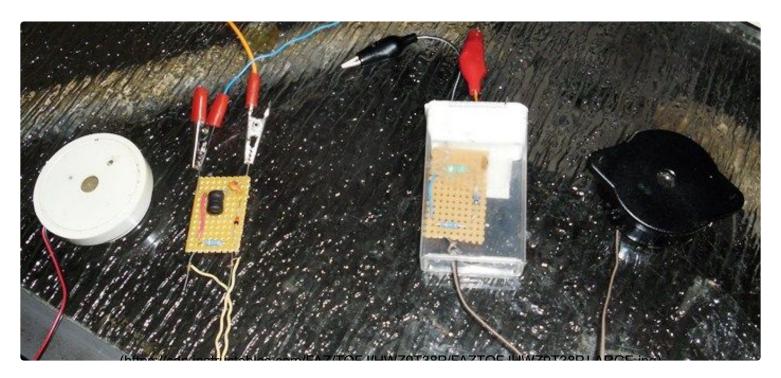
A metal rod or cold water pipe for the ground stake. Be careful which pipes you connect your Ground wire to.
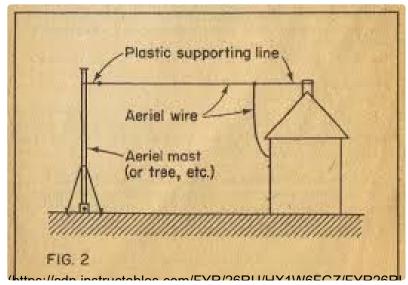
You will need a small sharp object for punching holes in the Tic Tac box.
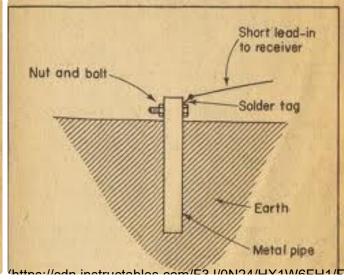
A small pair of wire cutters for cutting and stripping wires.

Add Tip   Ask Question

Step 2: How a Crystal Radio Works

Plastic supporting line

Aerial wire

Aerial mast
(or tree, etc.)

FIG. 2

Short lead-in
to receiver

Nut and bolt

Solder tag

Earth

Metal pipe

Radio signals consists of two parts - the 'carrier wave' which is the AM radio station's frequency of operation, and the 'program signal' which is mixed with the carrier wave for transmission.

Radio waves travel out from the AM transmitter tower through the atmosphere. We want to capture one specific frequency so we can listen to it, so we need the antenna/ground wire system to capture that signal. We also need a 'tuned circuit' that will filter out the desired AM signal, and discard the rest, so that all the other unwanted radio signals pass out through the ground wire to earth.

Two components in our circuit will perform that task for us. The capacitor C1, together with L1 inductor, form a basic 'series tuned' circuit. Their respective values will determine just which local AM radio station we will capture.

We also need a diode to 'detect' the voice and music, so we can hear them in our earphone, which transduces electrical signals into sound waves that we can hear.

In the photo above you can see a completed Tic Tac crystal radio. It is already inside the box. The other set is connected to the antenna/ground wire circuit, undergoing a 'soak test'. It is necessary to do this to ensure that the radio will work once inside the box!

The first diagram shows a typical antenna wire installation. Coming out of a window, the wire is anchored to the building and then over some distance (10 metres +) to a nearby tree or other building.

You must take great care not to erect antenna wires near to power or telephone cables, near your home!
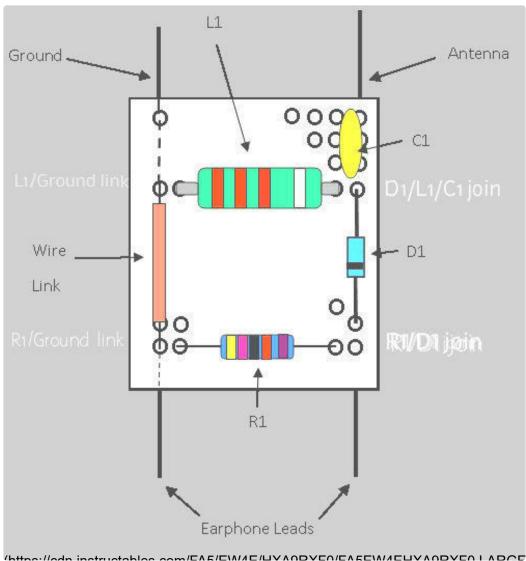
The ground wire comes out the same window and is anchored to a metal pipe/water pipe or metal ground stake, embedded in soft, moist soil.
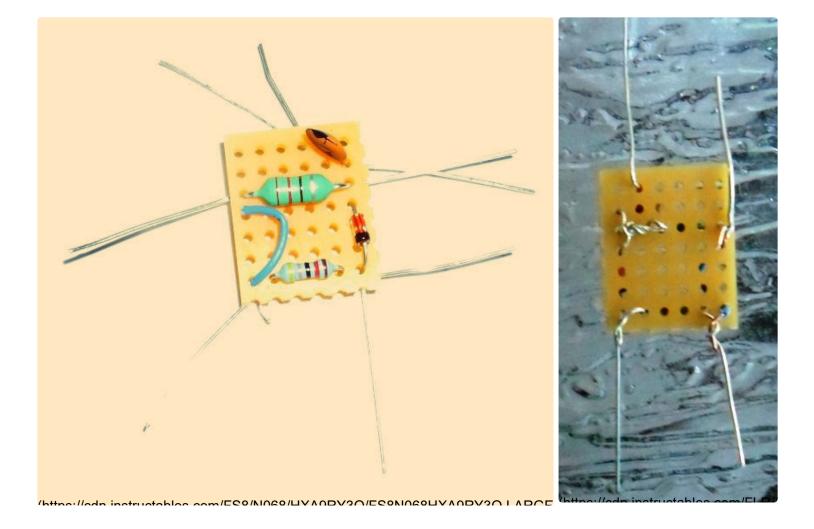
You must not connect ground wires to electrical mains wiring installations, including wall power outlets!

Another electrical hazard to consider is lightning strikes! Although it is very rare for anyone to be seriously injured or die from a lightning strike it is not impossible. So, if you hear a thunderstorm coming your way (you may hear the lightning 'crashes' in your earphone first,) then disconnect your antenna wire immediately, connect it to the ground wire and put it well up out of the way. Stay well clear of this temporary antenna/ground connection until the storm has completely passed away from your general area - miles away!

Add Tip   Ask Question
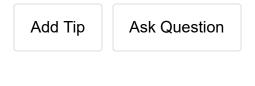
Step 3: How to Build the Tic Tac Crystal Radio Set

Ground

L1

Antenna

C1

L1/Ground link

D1/L1/C1 join

Wire

D1

Link

R1/Ground link

R1/D1 join

R1

Earphone Leads

(https://cdn.instructables.com/FA5/FW4F/HXA9BXF0/FA5FW4FHXA9BXF0.LARGE
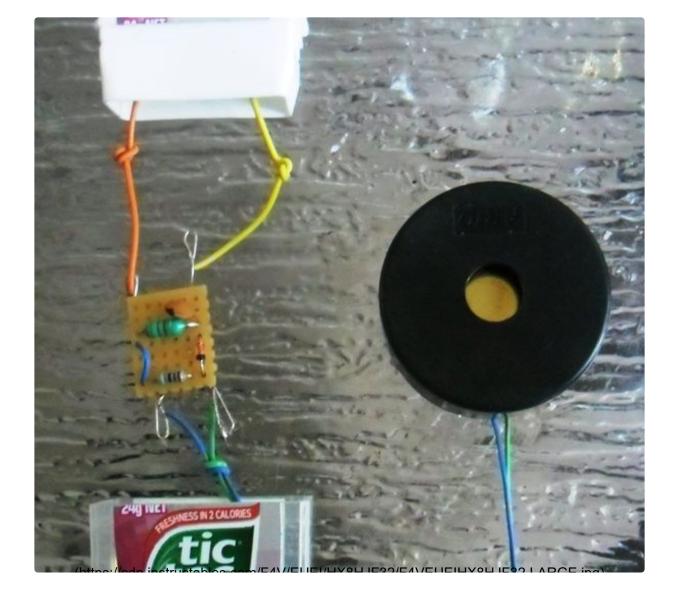
1. Take the components and lay them out on a clean surface.

2. Take the square of matrix board (this is also called 'perf board' because of all the holes or perforations in the material) and beginning with the 220uH (uH = micro-Henry) and place it as shown in the diagram.

3. Then take the diode, capacitor and resistor and place them in the places shown for them, making sure that the coloured band at one end of the diode joins with the resistor wire, as shown

4. Then take each junction where wires come together through their respective holes and gently twist them together, until they form a neat, tight bundle. Take your side cutters and cut off any excess length, taking care not to cut any one wire too short, so that it comes undone from the join.

.5. Take the link wire, strip 2 centimetres of insulation from each end of the wire and install that wire between the free ends of the inductor and resistor, and the matrix board construction is complete.

6. Then take the ceramic earphone, cut the plug off the end, and strip the 2 wire ends about 1.5 centimetres in length. Wrap each earphone wire around each end of the resistor component, underneath the matrix board.

7. Finally, strip the insulation off both ends of the 60mm long differently coloured wires, and attach them to the matrix board - one goes to the inductor/link wire junction and this will be the Ground connection wire. Attach the other one to the free end of capacitor C1 and this will be the antenna wire connection. Both connections are made underneath the matrix board. The Nic Nac Tic Tac Crystal Radio is now ready for testing.

Add Tip   Ask Question

Step 4: Final Checks and Installation of Wires

When you've finished constructing the matrix board circuit and have clipped all the excess component leads off, puncture three (3) holes in the Tic Tac box - 2 small holes on the top lid of the box, about an inch (24mm) apart, so that your antenna and ground wire leads can come through the box lid, and be connected to the matrix board at those 2 points.

Feed the wires through the holes and then tie small knots in each one, near the underside of the box lid, so that they won't pull back out if strained, and disconnect themselves from the matrix board.
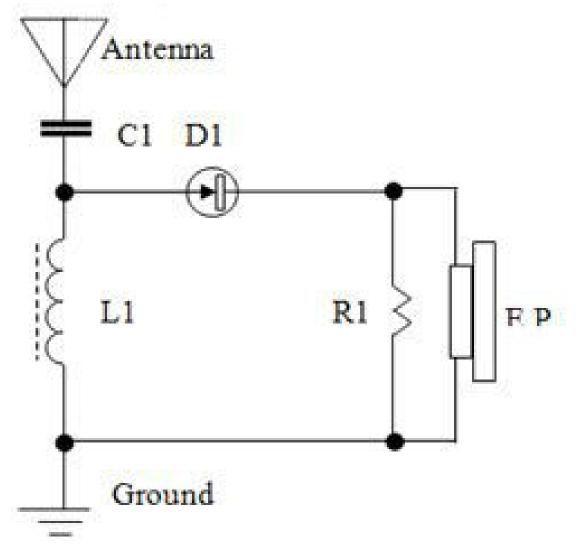
Then make one larger hole in the centre of the bottom of the box (clear part) so that your earphone wires can be fed through to the connecting points on that part of the matrix board. Tie a larger knot in the earphone lead so as to prevent it from pulling

out of the box if it is strained. Strip the insulation off the earphone leads, and wrap them around the matrix board leads at those 2 points. You are now ready to test out the Nic Nac Tic Tac Crystal Radio.

Add Tip   Ask Question

Step 5: Testing Out the Tic Tac Radio

(https://cdn.instructables.com/FX9H2HCHX8HJF34/FX9H2HCHX8HJF34.LARGE.j

Antenna

C1 D1

L1

R1

F P

Ground

Take your finished Tic Tac Radio (with the lid part gently pushed back inside the top of the box) and connect your Antenna and Ground wires to the lead outs from the box lid. Place the earphone in/over your ear,and listen carefully for a local AM radio station. This crystal radio is a simple one, and you may have to make one or two adjustments to the components, before you succeed in receiving one or more local AM radio stations, in your area.

If you can't hear anything in the earphone, don't panic. It might just be a simple wiring mistake, which is easily fixed. Go back over all of the steps, making sure that you have the right value components from the electronics store. Make sure that each component is in the right place on the matrix board, (don't confuse the L1 inductor with the R1 resistor - they look a lot like each other!) and that no component wires have come undone from the twisted joins you have just made. I'll

be writing up a troubleshooting step soon, so if you run into any problems, post your questions in the 'comments' section down at the bottom of the page, and I'll try and answer them as soon as possible.

Make sure that your antenna and ground wires aren't snagged on anything metal or anything dangerous!

If you have ANY doubts about the electrical safety of your antenna or ground wires, then consult a licensed electrical trades person, who will be able to advise you on electrical safety principles and procedures!

Always remember that electrical safety is your responsibility! If you don't think it is safe to proceed, then don't!
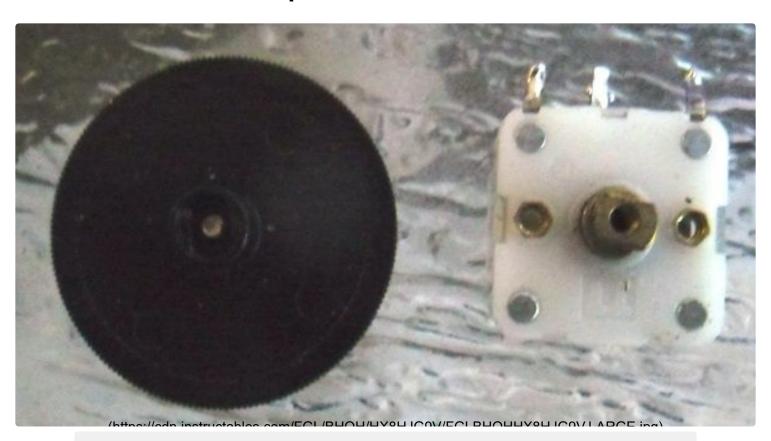
You will need at least 10 to 15 metres of antenna wire, strung between 2 insulating points (not connected to anything metallic, or that gets wet,) at least 2 to 3 metres in height - anything less than this minimum arrangement may mean that you cannot receive any signals at all.
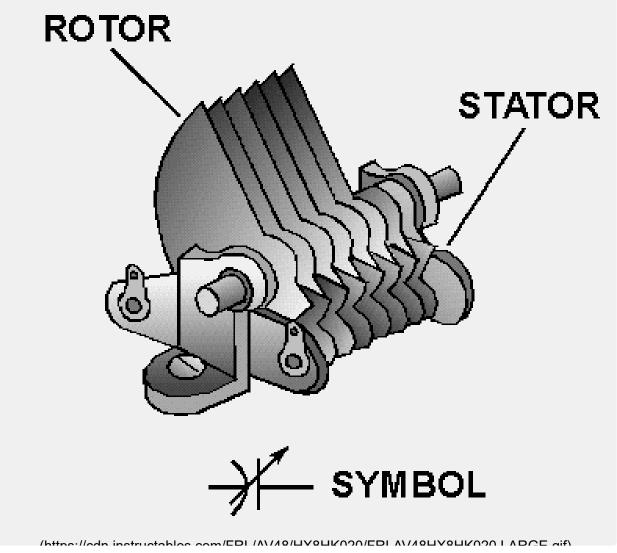
Some places are regarded as 'radio dead spots', so you may need to try an open space, such as a park or a remote corner of a beach. If you do erect antenna and ground wires in public places, hang some streamers or balloons off of them, to alert people to their presence, otherwise people going past may become entangled with them - and get cranky with you!!!

As a final word for now, you'll be happy to know that the completed "Nic Nac Tic Tac Radio" shown in the picture above, picked up local AM radio station 1116 khz 4BC here in Brisbane, with a very clear signal and quite good volume. It works! So be safe kids, have some fun and look forward to more... mk484

Add Tip   Ask Question

Step 6: Nic Nac Extras

(https://cdn.instructables.com/FCL/BHOH/HX8HJC9V/FCLBHOHHX8HJC9V.LARGE.jpg)

ROTOR

STATOR

SYMBOL

(https://cdn.instructables.com/FRLA/V48/HY8HK020/FRLAV48HX8HK020.LARGE.gif)

If you've built your Tic Tac crystal set and found that you can't receive a local AM radio station yet - don't panic - help is in on the way. You have just built the simplest version of the Tic Tac crystal radio and you may need to add one more component part for it to work properly. This is called a 'polyvaricon' - a miniature tuning capacitor, which can vary the frequency that your radio will receive at. You can see a picture of one up above - the small white box with the black knob next to it. It has 3 connecting tags - the one in the middle ('G') goes to the moving plates and the shaft, while the 2 outer tags ('O' and 'A') go to 2 sets of fixed plates. the smaller set of plates has a value of 60pF - pF is short for 'picoFarad' - a unit of measurement for capacitance. The larger set of plates is valued at 160 pF so that the combined value of the polyvaricon is 220 pF - or 220 picoFarads.

The other picture shows you what happens inside a basic tuning capacitor. there are 2 sets of metal plates - one set is fixed and the other set moves on a rotating shaft, connected to the tuning knob of your radio set. Both sets of plates are mounted on an insulating frame so that they won't 'short out' by touch each other.

The fixed capacitor C1, was chosen to tune somewhere close to the middle of the AM radio band. This band of frequencies starts at 531 kilohertz (Khz) and goes as far as 1701 Khz here in Australia. So we need a combination of coil and capacitor which will tune across all of those frequencies. Our simple Tic Tac radio is known as a 'series tuned' set. If you look at the circuit diagram, you can see electronic symbols for all components in the radio set.If you start at the top with the antenna symbol, you can see the capacitor C1 underneath that, the inductor L1 underneath C1 and then finally, the ground symbol - all wired in series with each other.
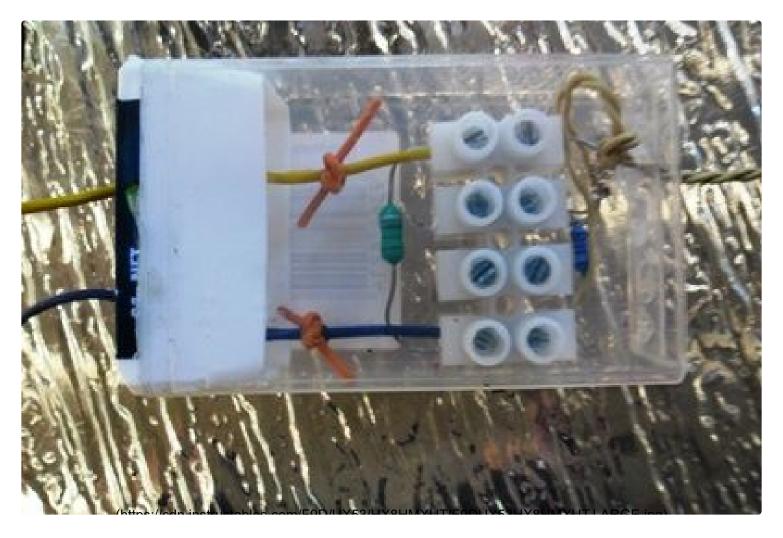
The diode D1 detector, the resistor and the earphone plus the link wire, can be considered as one block - the "detector unit". The Antenna wire, ground wire, inductor/coil and capacitor/polyvaricon, can be also be considered as another block - the "tuned circuit". So joining both blocks together, we have the tuned circuit that tunes in only one frequency, passing all other unwanted signals out through the ground wire to earth. This one 'tuned frequency' passes through the diode detector, which strips away the 'carrier wave' and leaves only the 'program signal' (voice,
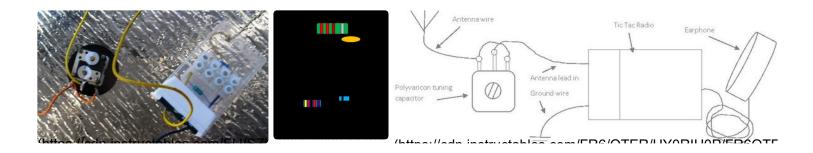
music etc,) behind, which is then fed via the resistor into the earphone. The earphone changes electrical impulses from the diode detector, into sound waves that we can hear. You need the R1 resistor to provide a pathway for the signals going through the diode, out to the ground wire connection. Without this resistor, the signals would sound very distorted and you couldn't hear the program signal very clearly.

Add Tip   Ask Question

Step 7: Adding the Polyvaricon Tuning Capacitor

If you want to expand the tuning range of your Tic Tac Radio, than all you have to do is a simple modification (change) to the circuit of your radio set. You can see from the 2 pictures above, that there's an alternative way of building the crystal radio - you can use the matrix board method or you can use a 4 way screw terminal strip.

To use the screw terminal version, cut yourself a 4 way strip of terminals as shown in the diagram and pictures. Undo the screws right out as far as they will go without falling out of the strip. Connect the component leads and wires from the antenna/ground system as well as the earphone wires. Cut off any excess from component leads that you don't need. Wrap thin wires from the antenna/ground system and the earphones, around the thicker component leads before screwing the screws in tight.
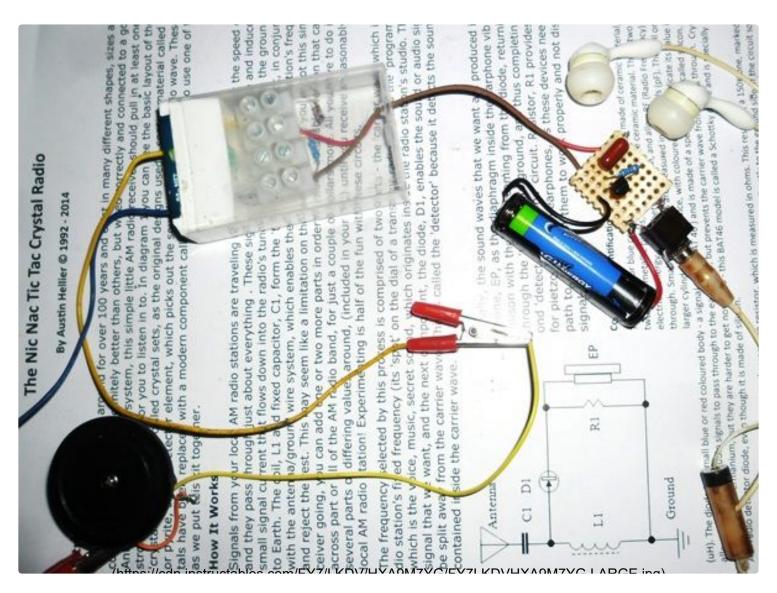
Regardless of which method you have used, all we will now wire up the tuning capacitor between the actual Antenna wire and the junction (join) where the diode D1, the fixed capacitor C1 and one end of the inductor meet. Remove the capacitor from the matrix board, and connect the Antenna lead out wire straight to the join of the diode and inductor. Then take the polyvaricon and another piece of wire. Strip both ends of that wire and join the two outer tags (tagged 'O' and 'A' - the centre on is tagged 'G') and then connect you actual Antenna wire to one of the outer tags. Connect the antenna lead out wire, coming out of the box, to the 'G' (middle) tag of the tuning capacitor,so that the antenna wiring now looks like the picture up above.

If you're having problems following the pictures, then refer to the diagrams, which clearly shows all of the connections Make sure your ground wire is connected to the set, and then, listening with your earphone in/over your ear, slowly turn the

tuning cap's flat knob, until you hear one or more stations. Congratulations - you now have a "tuneable" Tic Tac crystal set! Happy listening! And don't forget to post in your results, questions, problems etc... mk484 :)

Add Tip   Ask Question

Step 8: Tic Tac One Transistor Amplifier - Use Your Earbuds and Hear Great Sound...

Hi there folks - this isn't really an extra 'Step' as such - it's a sneak preview at a picture of my upcoming Instructable for a one transistor amplifier, which will connect to the Tic Tac Radio - and give you some really good volume - in your Iphone
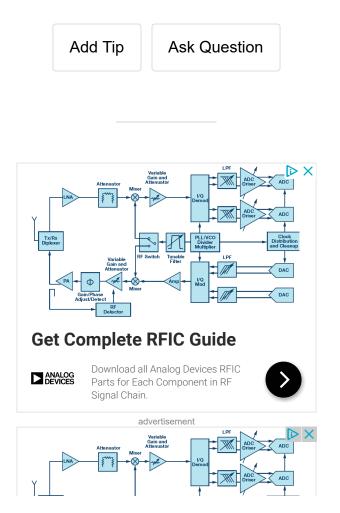
earbuds!

Yes, it's entirely possible nowadays, with modern circuit design, to fit a one transistor amp in such a small space (yes kiddies - it WILL fit inside the smaller Tic Tac box...) and at the same time, get that great sound that comes from those "inside your ear" type earbuds.
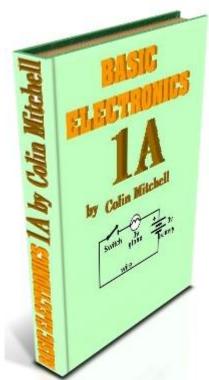
This circuit took me about a half hour to complete, uses only 3 electronic components, costing about $1.00, a stereo earphone jack and a 1.5 volt AAA battery and plastic holder costing about another $3 to $4, so you can build the amplifier for about $5 all up - don't forget to shop around and - mums and dads - cheap batteries will be OK for this project, and the battery can simply be replaced without soldering or undoing half a dozen screws....

I can hear my Tic Tac Radio with great volume and clarity - and so will you, so "stay tuned" (ha ha ha) to this series of Instructables, and you will hear great sound too...

mk484

Add Tip   Ask Question

go to: Talking Electronics Website

For any enquiries email **Colin Mitchell**

THE MULTIMETER

INDEX

TWO MULTIMETERS

There are basically two different types of MULTIMETER. ANALOGUE and DIGITAL
Analogue Multimeters have a NEEDLE or POINTER that moves across a scale.
Digital Multimeters have a numeric display of 3 or more digits. A Digital Multimeter with 3½ digits means the first digit shows only "1."

You really need both types to cover the number of tests needed for designing and repair-work. We will discuss how they work, how to use them and some of the differences between them.

The black (negative lead) ALWAYS stays in the

**black hole and the red lead changes to the
other red hole to measure 10 amps**

BUYING A MULTIMETER

There are many different types on the market.
The cost is determined by the number of ranges and also the extra features such as diode tester, buzzer (continuity), transistor tester, high DC current and others.
Since most multimeters are reliable and accurate, buy one with the greatest number of ranges at the lowest cost. The cheapest multimeters are on **eBay**.
This article explains the difference between an analogue meter and a digital meter.

Multimeters are sometimes called a "meter", a "VOM" (Volts-Ohms-Milliamps or Volt Ohm Meter) or "multi-tester" or even "a tester" - they are all the same.

One term used to describe a **DIGITAL MULTIMETER** is **3½ digits**.
This is the number of digits on the display. The first digit is usually made from two pixels and can only produce "1." This is called a half-digit. The other digits are full digits. The cheapest digital multimeters have 3½ digits. This will produce a reading of 1999 and the decimal point can produce values from 1.999 to 19.99 to 199.9 to 1999.
Another term is **DISPLAY COUNTS**. This is connected with the accuracy of the display, but since digital meters are accurate to 1% or less and we are using resistors with an accuracy of 5%, even a $10.00 digital meter will be perfect.

USING A MULTIMETER

Analogue and **digital** multimeters have either a rotary selector switch or push buttons to select the appropriate function and range. Some Digital Multimeters (DMMs) are auto ranging; they automatically select the correct range of voltage, resistance, or current when doing a test. However you need to select the function.

Before making any measurement you need to know what you are checking. If you are measuring voltage, select the AC range (10v, 50v, 250v, or 1000v) or DC range (0.5v, 2.5v, 10v, 50v, 250v, or 1000v). If you are measuring resistance, select the Ohms range (x1, x10, x100, x1k, x10k). If you are measuring current, select the appropriate current range DCmA 0.5mA, 50mA, 500mA, 10A. Every multimeter is different however the photo below shows a low cost Analogue multimeter with the basic ranges.

An ANALOGUE MULTIMETER

The most important point to remember is this:
You must select a voltage or current range that is bigger or HIGHER than the maximum expected value, so the needle does not swing across the scale and hit the "end stop."
If you are using a DMM (Digital Multi Meter), the meter will indicate if the voltage or current is higher than the selected scale, by showing "OL" - this means "Overload."
If you are measuring resistance such as 1M on the x10 range the "OL" means "Open Loop" and you will need to change the range. Some meters show "1' on the display when the measurement is higher than the display will indicate and some flash a set of digits to show over-voltage or over-current. A "-1" indicates the leads should be reversed for a "positive reading."
If it is an AUTO RANGING meter, it will automatically produce a reading, otherwise the selector switch must be changed to another range.

A typical DIGITAL Multimeter

**The Common (negative) lead ALWAYS fits into
the "COM" socket. The red lead fits into the
red socket for Voltage and Resistance.
Place the red lead (red banana plug)
into "A" (for HIGH CURRENT "Amps")
or mA,uA for LOW CURRENT.**

The black "test lead" plugs into the socket marked "-" "Common", or "Com," and the red "test lead" plugs into the meter socket marked "+" or "V-W-mA."   The third banana socket measures HIGH CURRENT and the positive (red lead) plugs into this. You DO NOT move the negative "-" lead at any time.
The following two photos show the test leads fitted to a digital meter. The probes and plugs have "guards" surrounding the probe tips and also the plugs so you can measure high voltages without getting near the voltage-source.

Analogue meters have an "Ohms Adjustment" to allow for the change in voltage of the battery inside the meter (as it gets old).

"Ohms Adjust" is also called "ZERO SET"
The sensitivity of this meter is 20,000ohms/volt
on the DC ranges and 5k/v on the AC ranges

Before taking a resistance reading (each time, for any of the Ohms scales) you need to "ZERO SET" the scale, by touching the two probes together and adjust the pot until the needle reads "0" (swings FULL SCALE). If the pointer does not reach full scale, the batteries need replacing. Digital multimeters do not need "zero adjustment."

ANALOGUE Vs DIGITAL

You cannot say one meter is better than the other because BOTH have advantages and disadvantages.
An analogue multimeter is the "old style" and it puts a load on a circuit and this may change the reading to give an incorrect readout, but it has the advantage of the needle moving across the scale fairly quickly so you can sometimes see if the voltage is fluctuating.
It also gives a more-accurate result in some high frequency circuits as it does not pick up stray fields and produce a false reading.
Digital meters put almost no load on a circuit and produce accurate readings from both low-impedance and high-impedance circuits.
Digital meters can display very low resistances.
You must remember to turn a Digital meter OFF to prevent the battery going flat.

If you are testing a circuit containing a high-frequency oscillator, use BOTH an ANALOGUE and DIGITAL meter to check the reading. Sometimes the leads of a Digital multimeter will pick up signals and create a false reading.
Sometimes you will get a voltage reading with a Digital multimeter due to a high resistance leak and a zero reading with an Analogue meter. This is why you need **BOTH** meters.

MEASURING VOLTAGE

Most of the readings taken with a multimeter will be VOLTAGE readings.
Before taking a reading, you should select the highest range and if the needle does not move up scale (to the right), you can select another range.
Always switch to the highest range before probing a circuit and keep your fingers away from the component being tested.
If the meter is Digital, select the highest range or use the auto-ranging feature, by selecting "V." The meter will automatically produce a result, even if the voltage is

AC or DC.

If the meter is not auto-ranging, you will have to select $V\!=\!=$ if the voltage is from a DC source or $V\!\sim$ if the voltage is from an AC source. DC means Direct Current (but this does not mean you select the CURRENT range - you are taking a voltage reading that is not rising and falling. That's why we say it is **DC** and do not say the words "direct current"). The voltage is coming from a battery or supply where it is steady and not "rising and falling."

You can measure the voltage at different points in a circuit by connecting the black probe to chassis. This is the 0v reference and is commonly called "Chassis" or "Earth" or "Ground" or "0v."

The red lead is called the "measuring lead" or "measuring probe" and it can measure voltages at any point in a circuit. Sometimes there are "test points" on a circuit and these are wires or loops designed to hold the tip of the red probe (or a red probe fitted with a mini clip).

You can also measure voltages ACROSS A COMPONENT. In other words, the reading is taken in PARALLEL with the component. It may be the voltage across a transistor, resistor, capacitor, diode or coil. In most cases this voltage will be less than the supply voltage.

If you are measuring the voltage in a circuit that has a [HIGH IMPEDANCE](), the reading will be inaccurate, up to 90% !!!, if you use a cheap analogue meter.

Here's a simple case.

The circuit below consists of two 1M resistors in series. The voltage at the mid point will be 5v when nothing is connected to the mid point. But if we use a cheap analogue multimeter set to 10v, the resistance of the meter will be about 100k, if the meter has a sensitivity of 10k/v and the reading will be incorrect.

Here how it works:

Every meter has a sensitivity. The sensitivity of the meter is the sensitivity of the movement and is the amount of current required to deflect the needle FULL SCALE. This current is very small, normally 1/10th of a milliamp and corresponds to a sensitivity of 10k/volt (or 1/30th mA, for a sensitivity of 30k/v).

If an analogue meter is set to 10v, the internal resistance of the meter will be 100k for a 10k/v movement.

If this multimeter is used to test the following circuit, the reading will be inaccurate.

The reading should be 5v as show in diagram **A**.

But the analogue multimeter has an internal resistance of 100k and it creates a circuit shown in **C**.

The top 1M and 100k from the meter create a combined PARALLEL resistance of 90k. This forms a series circuit with the lower 1M and the meter will read less than 1v

If we measure the voltage across the lower 1M, the 100k meter will form a value of resistance with the lower 1M and it will read less than 1v

If the multimeter is 30k/v, the readings will be 2v. See how easy it is to get a totally inaccurate reading.

A   **B**   **C**

D   **E**

If the reading is taken with a Digital Meter, it will be more accurate as a DMM does not take any current from the circuit (to activate the meter). In other words it has a very HIGH input impedance. Most Digital Multimeters have a fixed input resistance (impedance) of 10M - no matter what scale is selected. That's the reason for choosing a DMM for high impedance circuits. It also gives a reading that is accurate to about 1%.

to Index

MEASURING VOLTAGES in a CIRCUIT

You can take many voltage-measurements in a circuit. You can measure "across" a component, or between any point in a circuit and either the positive rail or earth rail (0v rail). In the following circuit, the 5 most important voltage-measurements are shown. Voltage "A" is across the electret microphone. It should be between 20mV and 500mV. Voltage "B" should be about 0.6v. Voltage "C" should be about half-rail voltage. This allows the transistor to amplify both the positive and negative parts of the waveform. Voltage "D" should be about 1-3v. Voltage "E" should be the battery

voltage of 12v.

MEASURING CURRENT

You will rarely need to take current measurements, however most multimeters have DC current ranges such as 0.5mA, 50mA, 500mA and 10Amp (via the extra banana socket) and some meters have AC current ranges. Measuring the current of a circuit will tell you a lot of things. If you know the normal current, a high or low current can let you know if the circuit is overloaded or not fully operational.

Current is always measured when the circuit is working (i.e: with power applied).
It is measured IN SERIES with the circuit or component under test.
The easiest way to measure current is to remove the fuse and take a reading across the fuse-holder. Or remove one lead of the battery or turn the project off, and measure across the switch.
If this is not possible, you will need to remove one end of a component and measure with the two probes in the "opening."
Resistors are the easiest things to desolder, but you may have to cut a track in some circuits. You have to get an "opening" so that a current reading can be taken.
The following diagrams show how to connect the probes to take a CURRENT reading.
Do not measure the current ACROSS a component as this will create a "short-circuit."
The component is designed to drop a certain voltage and when you place the probes across this component, you are effectively adding a "link" or "jumper" and the voltage at the left-side of the component will appear on the right-side. This voltage may be too high for the circuit being supplied and the result will be damage.

Measuring the current of a globe

Do NOT measure the CURRENT of a battery
(by placing the meter directly across the terminals)
A battery will deliver a very HIGH current
and damage the meter

Do not measure the "current a battery will deliver" by placing the probes across the terminals. It will deliver a very high current and damage the meter instantly. There are special battery testing instruments for this purpose.

When measuring across an "opening" or "cut," place the red probe on the wire that supplies the voltage (and current) and the black probe on the other wire. This will produce a "POSITIVE" reading.
A positive reading is an UPSCALE READING and the pointer will move across the scale - to the right. A "NEGATIVE READING" will make the pointer hit the "STOP" at the left of the scale and you will not get a reading. If you are using a Digital Meter, a negative sign "-" will appear on the screen to indicate the probes are around the wrong way. No damage will be caused. It just indicates the probes are connected incorrectly.
If you want an accurate CURRENT MEASUREMENT, use a digital meter.

MEASURING RESISTANCE

Turn a circuit off before measuring resistance.
If any voltage is present, the value of resistance will be incorrect.
In most cases you cannot measure a component while it is in-circuit. This is because the meter is actually measuring a voltage across a component and calling it a "resistance."
The voltage comes from the battery inside the meter. If any other voltage is present, the meter will produce a false reading.
If you are measuring the resistance of a component while still "in circuit," (with the power off) the reading will be lower than the true reading.

Measuring resistance

**Measuring resistance of a heater
(via the leads)**

Measuring the resistance of a piece of resistance-wire

Measuring the resistance of a resistor

Do not measure the "Resistance of a Battery"

1. Do not measure the "resistance of a battery." The resistance of a battery (called the Internal impedance) is not measured as shown in the diagrams above. It is measured by creating a current-flow and measuring the voltage across the battery. Placing a multimeter set to **resistance** (across a battery) will destroy the meter.
2. Do not try to measure the resistance of any voltage or any "supply."

Resistance is measured in OHMs.
The resistance of a 1cm x 1cm bar, one metre long is 1 ohm.
If the bar is thinner, the resistance is higher. If the bar is longer, the resistance is higher.
If the material of the bar is changed, the resistance is higher.
When carbon is mixed with other elements, its resistance increases and this knowledge is used to make RESISTORS.
Resistors have RESISTANCE and the main purpose of a resistor is to reduce the CURRENT FLOW.
It's a bit like standing on a hose. The flow reduces.
When current flow is reduced, the output voltage is also reduced and that why the water does not spray up so high. Resistors are simple devices but they produce many different effects in a circuit.
A resistor of nearly pure carbon may be 1 ohm, but when non-conducting "impurities" are added, the same-size resistor may be 100 ohms, 1,000 ohms or 1 million ohms.
Circuits use values of less than 1 ohm to more than 22 million ohms.

Resistors are identified on a circuit with numbers and letters to show the exact value of resistance - such as 1k   2k2   4M7

The letter Ω (omega - a Greek symbol) is used to identify (or express) (or represent) the word "Ohm."
But this symbol is not available on some word-processors, so the letter "R" is used. The letter "E" is also sometimes used and both mean "Ohms."
A one-ohm resistor is written "1R" or "1E." It can also be written "1R0" or "1E0."
A resistor of one-tenth of an ohm is written "0R1" or "0E1." The letter takes the place of the decimal point.
10 ohms = 10R
100 ohms = 100R

1,000 ohms = 1k (k= kilo = one thousand)
10,000 ohms = 10k
100,000 ohms = 100k
1,000,000 ohms = 1M   (M = MEG = one million)

The size of a resistor has nothing to do with its resistance. The size determines the wattage of the resistor - how much heat it can dissipate without getting too hot.
Every resistor is identified by colour bands on the body, but when the resistor is a surface-mount device, numbers are used and sometimes letters.
You MUST learn the colour code for resistors and the following table shows all the colours for the most common resistors from 1/10th of an ohm to 22 Meg ohms for resistors with 5% and 10% tolerance.

If 3rd band is gold, Divide by 10
If 3rd band is silver, Divide by 100
(to get 0.22ohms etc)

COLOR CODES FOR THE WHOLE E12/E24 RANGE OF RESISTORS
The twelve odd rows - 1, 3, 5...- represent values available in the E12 range only, plus 10M

to Index

MAKE YOUR OWN RESISTOR

Make your own variable resistor that changes resistance according to the pressure.
Use a piece of conductive foam used to package Integrated Circuits. You can ask at an electronics shop.
Use two coins or pieces of printed circuit board or aluminium foil for the top and bottom conductors.
You can solder wires to the PC board or fold the aluminium foil over a few times to hold the wires.
The resistance of the foam will reduce as you press on the "cell."
The actual resistance-values will depend on the size of the foam, the thickness and
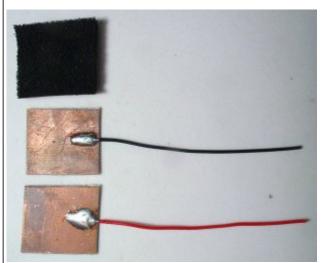
pressure.
This cell is a very simple cell called a **LOAD CELL**.
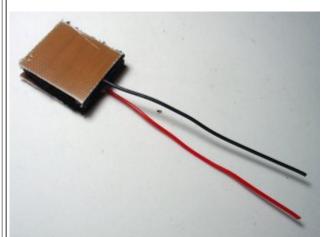
The top and bottom "plates"

The foam is placed between the plates.

The complete LOAD CELL

The resistance of the unloaded LOAD CELL

The fully loaded resistance can be as low as 9,330 ohms

to Index

MEASURING CONTINUITY

CONTINUITY is the same as **ZERO OHMS** or the resistance of a short length of wire. It can also mean the resistance through a switch or globe or a low-value resistor.
It basically means a "PATH" and sometimes refers to a whole circuit when the switch is closed. In other words CONTINUITY means we have a "circuit." We have "current flowing" and generally refers to a low-resistance circuit.
Both ANALOGUE and DIGITAL multimeters can measure CONTINUITY and you have to work out the approximate value of resistance for the circuit you are testing, - BEFORE TAKING A READING.
If the reading is above 300 ohms or contains a diode, you cannot use a DIGITAL MULTIMETER as the buzzer on the continuity setting will not respond.

The project being tested must not have the power applied as the resistance ranges on a multimeter are actually measuring a voltage across the leads and any voltage on the circuit or contained in any electrolytics, will upset the reading.

To take a reading with an **ANALOGUE** multimeter, select the x1 setting and the pointer will move across the scale to the actual value of resistance.
It it move full scale, you have ZERO OHMS resistance and this can mean a short-circuit or continuity via a wire.
If a diode is in the circuit you must also reverse the leads to get a reading.
The resistance of a globe will be very low when it is not illuminated, so don't think a fault is present.
Measuring CONTINUITY is the same as measuring LOW RESISTANCE.

To take a reading with a **DIGITAL** multimeter, select the buzzer setting. It will respond if the resistance is less than 300 ohms.   It will not respond if a diode is in the circuit.

Meter set to BUZZER - CONTINUITY

You can also use the x1 resistance setting to get an accurate value of resistance. Touch the probes together to get the initial reading and subtract this value from the final reading.
When probing a circuit containing electrolytics, you may get a beep from the buzzer. This indicates the resistance is low because the multimeter is charging the electrolytic and it will beep until the electrolytic is charged to about 0.7v.
The same applies when probing across the power rails of a circuit. The circuit may contain electrolytics that will charge when probing and the buzzer will beep.
The Digital multimeter is actually detecting a voltage less than 0.7v across the probes and is created by a voltage-divider network inside the meter.
The voltage divider put 2v across the probes and when this drops to less than 0.5v, the buzzer is activated. That why it odes not buzz when testing a diode as the diode drops the voltage to 0.6v.

MEASURING A DIODE
A diode can be measured to see if it is "open" or "damaged" or "working" by placing the probes across the component.
If the diode is "open" (it will not work), the needle will NOT swing across the scale when touching the component with the probes in one direction or when the probes are reversed.
If the diode is "damaged" (does not work), the needle will swing fully across the scale when touching the component with the probes in one direction or when the probes are reversed.
If the diode is FUNCTIONAL, (works) the needle will swing about mid-way when touching the leads of the diode in one direction and it will not move when the probes are reversed.

WHY?
The positive of the battery inside an analogue multimeter comes out the black probe and that is why you will get a reading when the probers are "around the wrong way." The needle will swing a different amount for each resistance setting on the dial as the needle represents 0.6v drop and NOT an actual resistance.

There are two things you must remember.
1. When the diode is measured in one direction, the needle will will **not move at**

all. The technical term for this is the diode is **reverse biased**. It will not allow any current to flow. Thus the needle will not move.
When the diode is connected around the other way, the needle will swing to the right (move up scale) to about 80% of the scale. This position represents the voltage drop across the junction of the diode and is NOT a resistance value. If you change the resistance range, the needle will move to a slightly different position due to the resistances inside the meter. The technical term for this is the diode is **forward biased**. This indicates the diode is not faulty.
The needle will swing to a slightly different position for a "normal diode" compared to a Schottky diode. This is due to the different junction voltage drops.
However we are only testing the diode at very low voltage and it may break-down when fitted to a circuit due to a higher voltage being present or due to a high current flowing.

2. The leads of an **Analogue Multimeter** have the positive of the battery connected to the black probe and the readings of a "good diode" are shown in the following two diagrams:

The diode is REVERSE BIASED in the diagram above and diodes not conduct.

The diode is FORWARD BIASED in the diagram above and it conducts

TESTING A DIODE ON A DIGITAL METER

A Digital multimeter will measure the voltage-drop across the diode when the probes are connected in one direction (approx 0.640 on the scale) and a high reading (1) in the other direction. You need to select the "DIODE" setting on the dial as the other settings will produce a meaningless reading.

cathode (k)

cathode (k)
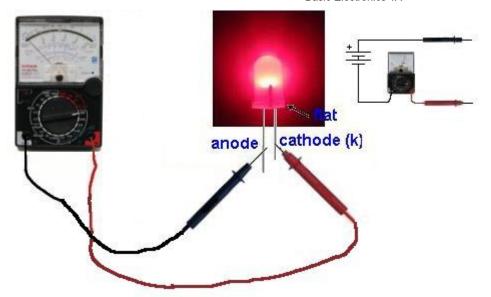
Some DIGITAL MULTIMETERS will show mV drop across the diode when the setting on the meter is "diode" or the "x1" or "x10" resistance range.

TESTING A LED
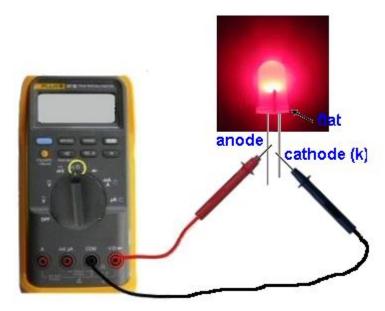
Some multimeters will test LEDs.
It depends on the voltage of the battery inside the case of the multimeter.
Many analogue multimeters have a single 1.5v cell and these cannot test LEDs.
Analogue Multimeters with 3v (for the resistance ranges) can test some LEDs.
White LEDs need about 3.6v and they may not illuminate on 3v.

**The negative lead of an ANALOGUE meter is POSITIVE!
The multimeter must have 3v (2 cells)**

Digital multimeters have a 9v battery and they will illuminate all colour LEDs when the leads are placed as shown in the diagram:

**A Digital meter will illuminate all LEDs and the black probe
touches the cathode.**

to Index

TESTING A TRANSISTOR WITH A DIGITAL METER

Testing a transistor with a **Digital Meter** must be done on the "DIODE" setting as a digital meter does not deliver a current through the probes on some of the resistance settings and will not produce an accurate reading.
The "DIODE" setting must be used for diodes and transistors. It should also be called a "TRANSISTOR" setting.

TESTING A TRANSISTOR WITH AN ANALOGUE METER

The first thing you may want to do is test an unknown transistor for COLLECTOR, BASE AND EMITTER. You also want to perform a test to find out if it is NPN or PNP. That's what this test will provide.
You need a cheap multimeter called an ANALOGUE METER - a multimeter with a

scale and pointer (needle).
It will measure resistance values (normally used to test resistors) - (you can also test other components) and Voltage and Current. We use the resistance settings. It may have ranges such as "x10" "x100"   "x1k"   "x10"
Look at the resistance scale on the meter. It will be the top scale.
The scale starts at zero on the right and the high values are on the left. This is opposite to all the other scales.
When the two probes are touched together, the needle swings FULL SCALE and reads "ZERO." Adjust the pot on the side of the meter to make the pointer read exactly zero.

How to read: "x10" "x100" "x1k"   "x10"
Up-scale from the zero mark is "1"
When the needle swings to this position on the "x10" setting, the value is 10 ohms.
When the needle swings to "1" on the "x100" setting, the value is 100 ohms.
When the needle swings to "1" on the "x1k" setting, the value is 1,000 ohms = 1k.
When the needle swings to "1" on the "x10k" setting, the value is 10,000 ohms = 10k.
Use this to work out all the other values on the scale.
Resistance values get very close-together (and very inaccurate) at the high end of the scale. [This is just a point to note and does not affect testing a transistor.]

Step 1   - FINDING THE BASE and determining NPN or PNP
Get an unknown transistor and test it with a multimeter set to "x10"
Try the 6 combinations and when you have the black probe on a pin and the red probe touches the other pins and the meter swings nearly full scale, you have an NPN transistor. The black probe is BASE
If the red probe touches a pin and the black probe produces a swing on the other two pins, you have a PNP transistor. The red probe is BASE
If the needle swings FULL SCALE or if it swings for more than 2 readings, the transistor is **FAULTY**.

This is an NPN transistor
The black probe is the BASE

This is a PNP transistor
The red probe is the BASE

Step 2   - FINDING THE COLLECTOR and EMITTER
Set the meter to "x10k."
For an NPN transistor, place the leads on the transistor and when you press hard on the two leads shown in the diagram below, the needle will swing almost full scale.

Set to "x10k"

This is the
POSITIVE
of the battery
in the meter

push
HERE

This is an
NPN
transistor

This is the NEGATIVE
of the battery in the meter

Here is the equivalent circuit:

+

The finger is
"turning ON
the transistor"

collector

base

emitter

You have found the Collector Base and Emitter

For a PNP transistor, set the meter to "x10k" place the leads on the transistor and when you press hard on the two leads shown in the diagram below, the needle will swing almost full scale.

Set to "x10k"

This is the
POSITIVE
of the battery
in the meter

This is a
PNP
transistor

This is the NEGATIVE
of the battery in the meter

push
HERE

Here is the equivalent circuit:

The finger is "turning ON the transistor"

emitter

base

collector

You have found the Collector Base and Emitter

For more details on testing components with a multimeter, see:
Testing Electronic Components.

MULTIMETER TEST

1. What is the reading on the multimeter:

12V

Answer

2. What is the reading on the multimeter:

Answer

3. What is the reading on the multimeter:

Answer

4. What is the voltage between the following points:

Terminal block

| ⊗ 1 ⊗ |
| ⊗ 2 ⊗ |
| ⊗ 3 ⊗ |
| ⊗ 4 ⊗ |

Terminal block

| ⊗ 5 ⊗ |
| ⊗ 6 ⊗ |
| ⊗ 7 ⊗ |
| ⊗ 8 ⊗ |

Battery
(9 volts)

Light bulb

Between 1 & 3   Between 5 & 7   Between 3 & 8   Between 1 & 7

Answer   **Answer**   **Answer**   **Answer**

5. What is the voltage on each multimeter:

B

A

switch *closed*

9v

C

Meter A   **Meter B**   **Meter C**

Answer   **Answer**   **Answer**

6. What is the voltage on each multimeter:

B

A

C

switch *open*

9v

Meter A

Answer

Meter B

Answer

Meter C

Answer

7. What is the reading on the multimeter:

Ω

ᴬᵥV   ᴬᵥA

═V   ═A

OFF

V Ω

A   COM

Switch *ON*

SW1

C1

9v

Answer

8. The multimeter has three resistance ranges: **x1 ohms   x100 ohms   x10k ohms**.
How far across the scale will the needle move on each range for a 4k7 resistor:

x1   **x100**   **x10k**

Answer   **Answer**   **Answer**

9. The resistance ranges on an ANALOGUE multimeter use the battery inside the case to move
the pointer. If the multimeter is left on "**ohms range**" with the probes apart, will the battery go
flat?

Answer
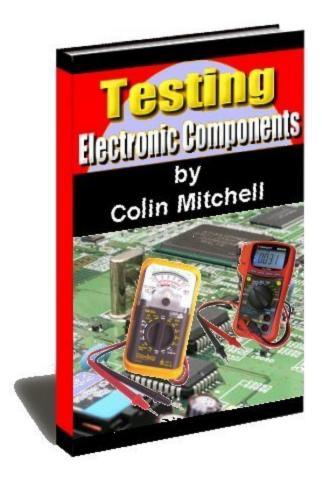
This eBook shows you how to TEST COMPONENTS.
To do this you need "TEST GEAR."   The best item of Test Gear is a
MULTIMETER. It can test almost 90% of all components. And that's what we
will do in this eBook:

CONTENTS

Use your "brain, knowledge and your fingers."

Before we start, fixing anything is a combination of skill, luck and good diagnosis.
Sometimes you can fix something by letting it run until it finally fails.
Some things start to work as soon as you touch them.
Some things can never be fixed.
But some things can be fixed by feeling the temperature rise and deciding if it is getting too hot.
Sometimes you can smell something getting too hot.
Sometimes you can see SMOKE.
All these things make you a very clever technician and about 50% of faults will be fixed by looking for dry joints, burnt parts, overheating and carefully inspecting an item before you disturb it.

By simply touching different items you can quite often feel a hot item and home-in on the fault - at a saving of hours of work.
Servicing is not "A bull at a gate" approach.
You may be able to service something by turning it on and leaving it for hours - and start thinking.
It may take you a day to come up with the answer.
Believe me. That's how it worked for me - while fixing over 35,000 TV's.

TEST EQUIPMENT

Everyone thinks TEST EQUIPMENT will "solve the problem."
This is a big big MISTAKE.
Test equipment can help solve a problem and it can "lead to frustration," "give an incorrect answer," "mess you up," and make things worse.
You have to be very careful with test equipment and especially EXPENSIVE equipment because it is very sensitive and can detect pulses and glitches and voltages that are not affecting the operation of the circuit.
You will learn a lot of tricks when reading through this article, but let me say two things.
There are lots of faults and components that you cannot test with "test equipment" because they are either intermittent or the equipment does not load the device to the same extent as the circuit.
And secondly you need both an ANALOGUE multimeter and a DIGITAL meter to cover all the situations.
And if you are working on a car, you only need a $5.00 analogue meter because it will be dropped or fall into a crack, and you will only lose $5.00
You will learn that a digital meter will pick up spikes and signals on a line and show an incorrect reading.
That's why you need to back-up your readings with an analogue meter.
When you charge a battery it gets a "floating voltage" and this will be higher than the actual voltage, when the battery is fitted to a project. An analogue meter will draw a slight current and remove the "floating voltage."
Component testers can also give you a false reading, either because the component is out of range of the tester or intermittent and you need to be aware of this.
Oscilloscopes can also display waveforms that are parts of glitches or noise from other chips and these do not affect the operation of the part of the circuit you are investigating.
Sometimes you cannot pickup a pulse because it is not regular and the trigger on the oscilloscope does not show it on the screen. You may think it is missing.
It all depends on the "speed of the oscilloscope" - it's maximum frequency of operation.
Lastly- Power Supplies. You cannot test globes and motors on a power supply because the starting current can be 5 times more than the operating current. The power supply may not be able to deliver this high current and thus you will think the motor or globe is faulty.

MULTIMETERS

There are two types:
DIGITAL and ANALOGUE
A **Digital Multimeter** has a set of digits on the display and an Analogue Multimeter has a scale with a pointer (or needle).
You really need both types to cover the number of tests needed for designing and repair-work. We will discuss how they work, how to use them and some of the differences between them.

DIGITAL AND ANALOGUE MULTIMETERS

BUYING A MULTIMETER

There are many different types on the market.
The cost is determined by the number of ranges and also the extra features such as diode tester, buzzer (continuity), transistor tester, high DC current and others.
Since most multimeters are reliable and accurate, buy one with the greatest number of ranges at the lowest cost.
This article explains the difference between a cheap analogue meter, an expensive analogue meter and a digital meter. You will then be able to work out which two meters you should buy.

Multimeters are sometimes called a "meter", a "VOM" (Volts-Ohms-Milliamps or Volt Ohm Meter) or "multi-tester" or even "a tester" - they are all the same.

USING A MULTIMETER

Analogue and digital multimeters have either a rotary selector switch or push buttons to select the appropriate function and range. Some Digital Multimeters (DMMs) are auto ranging; they automatically select the correct range of voltage, resistance, or current when doing a test. However you need to select the function.

Before making any measurement you need to know what you are checking. If you are measuring voltage, select the AC range (10v, 50v, 250v, or 1000v) or DC range (0.5v, 2.5v, 10v, 50v, 250v, or 1000v). If you are measuring resistance, select the Ohms range (x1, x10, x100, x1k, x10k). If you are measuring current, select the appropriate current range DCmA 0.5mA, 50mA, 500mA. Every multimeter is different however the photo below shows a low cost meter with the basic ranges.

The most important point to remember is this:
You must select a voltage or current range that is bigger or HIGHER than the maximum expected value, so the needle does not swing across the scale and hit the "end stop."
If you are using a DMM (Digital Multi Meter), the meter will indicate if the voltage or current is higher than the selected scale, by showing "OL" - this means "Overload." If you are measuring resistance such as 1M on the x10 range the "OL" means "Open Loop" and you will need to change the range. Some meters show "1' on the display when the measurement is higher than the display will indicate and some flash a set of digits to show over-voltage or over-current. A "-1" indicates the leads should be reversed for a "positive reading."
If it is an AUTO RANGING meter, it will automatically produce a reading, otherwise the selector switch must be changed to another range.

**The Common (negative) lead ALWAYS fits into
the "COM" socket. The red lead fits into the
red socket for Voltage and Resistance.
Place the red lead (red banana plug)
into "A" (for HIGH CURRENT "Amps")
or mA,uA for LOW CURRENT.**
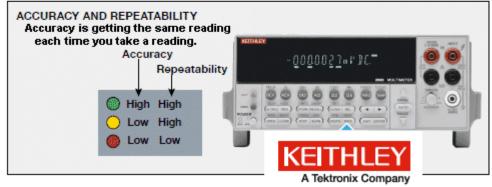
The black "test lead" plugs into the socket marked "-" "Common", or "Com," and
the red "test lead" plugs into meter socket marked "+" or "V-W-mA."   The third
banana socket measures HIGH CURRENT and the positive (red lead) plugs into this.
You DO NOT move the negative "-" lead at any time.
The following two photos show the test leads fitted to a digital meter. The probes
and plugs have "guards" surrounding the probe tips and also the plugs so you can
measure high voltages without getting near the voltage-source.

What's the difference between **sensitivity, resolution,** and **accuracy?**

RESOLUTION

The smallest portion of the signal that can be measured and displayed on any range. The resolution on this Model DMM7510 7.5-digit DMM, 100mV range is 10nV. On the 1V range it is 100nV.

-010.00004 mV

SENSITIVITY

The smallest change in the input signal that can be detected. On the 100mV range, this DMM is capable of sensitivity down to tens of nanovolts (nV).

Local   defbuffer1   No Script   CONT
MEASURE DC VOLTAGE   AUTO
AZERO
Range   Auto   REL
100mV
FUNCTIONS

DCV	DCI	2W Ω	Freq	Temp
ACV	ACI	4W Ω	Period	Cap
Digi V	Digi I	Cont	Diode	Ratio

ACCURACY AND REPEATABILITY
Accuracy is getting the same reading each time you take a reading.

Accuracy
Repeatability

-0000027 nF DC

🟢 High High
🟡 Low High
🔴 Low Low

KEITHLEY
A Tektronix Company

The question above applies to both (every) type of multimeter and the type of meter you use depends on the accuracy you need. Sometimes you are looking for 1mV change on a 20v rail. Only a DMM will (or a CRO) will produce a result.

Analogue meters have an "Ohms Adjustment" to allow for the change in voltage of the battery inside the meter (as it gets old).

"Ohms Adjust" is also called "ZERO SET"
The sensitivity of this meter is 20,000ohms/volt
on the DC ranges and 5k/v on the AC ranges

Before taking a resistance reading (each time on any of the Ohms scales) you need to "ZERO SET" the scale, by touching the two probes together and adjust the pot until the needle reads "0" (swings FULL SCALE). If the pointer does not reach full

scale, the batteries need replacing. Digital multimeters do not need "zero adjustment."

FIXING A MULTIMETER

A multimeter can get "broken" "damaged" and go "faulty."
I don't know why, but eventually they stop working.
It can be something simple like a flat battery, corroded battery contacts, broken switch or something complex, like the circuitry failing.
Multimeters are so cheap, you can buy a new one for less than $10.00
These meters can have a 10 amp range, transistor tester and measure up to 2 meg ohms.
That's why I suggest buying a $10.00 meter. They are just as good as a $60.00 meter and the cheapest meters last the longest.
Dropping an analogue meter can cause the hair spring to loop over one of the supports and the needle will not zero correctly. You will need to open the cover on the movement and lift the spring off the support with a needle.
A faulty meter can be used in a battery-charger circuit to measure the current or voltage if that scale is still reading-correctly.
Otherwise keep the leads and throw the meter out. It is too dangerous keeping a meter that shows an incorrect reading.

MEASURING FREQUENCY

Before we cover the normal uses for a multimeter, it is interesting to note that some **Digital Multimeters (DMM)** have features such as Capacitance, Frequency and measuring the gain of a transistor as well as a number of other features using probes such as a temperature probe. The VICHY VC99 meter above is an example and costs about $40.00.

Basic function	Range
DCV	600mV/6V/60V/600V/1000V
ACV	6V/60/600/1000V
DCA	600uA/6000uA/60mA/600mA/6A/20A
ACA	600uA/6000uA/60mA/600mA/6A/20A
Capacitance	40nF/400nF/4uF/40uF/400uF/4000uF
Resistance	600Ω/6kΩ/60kΩ/600kΩ/6MΩ/60MΩ
Frequency	100Hz/1kHz/10kHz/100kHz/1MHz/60MHz

Temperature	-40°C~1000°C
	0°F~1832°F

MEASURING VOLTAGE

Most of the readings you will take with a multimeter will be VOLTAGE readings.
Before taking a reading, you should select the highest range and if the needle does not move up scale (to the right), you can select another range.
Always switch to the highest range before probing a circuit and keep your fingers away from the component being tested.
If the meter is Digital, select the highest range or use the auto-ranging feature, by selecting "V." The meter will automatically produce a result, even if the voltage is AC or DC.
If the meter is not auto-ranging, you will have to select $V =$ if the voltage is from a DC source or $V \sim$ if the voltage is from an AC source. DC means Direct Current and the voltage is coming from a battery or supply where the voltage is steady and not changing and AC means Alternating Current where the voltage is coming from a voltage that is rising and falling.
You can measure the voltage at different points in a circuit by connecting the black probe to chassis. This is the 0v reference and is commonly called "Chassis" or "Earth" or "Ground" or "0v."
The red lead is called the "measuring lead" or "measuring probe" and it can measure voltages at any point in a circuit. Sometimes there are "test points" on a circuit and these are wires or loops designed to hold the tip of the red probe (or a red probe fitted with a mini clip or mini alligator clip).
You can also measure voltages ACROSS A COMPONENT. In other words, the reading is taken in PARALLEL with the component. It may be the voltage across a transistor, resistor, capacitor, diode or coil. In most cases this voltage will be less than the supply voltage.
If you are measuring the voltage in a circuit that has a HIGH IMPEDANCE, the reading will be inaccurate, up to 90% !!!, if you use a cheap analogue meter.

Here's a simple case.

The circuit below consists of two 1M resistors in series. The voltage at the mid point will be 5v when nothing is connected to the mid point. But if we use a cheap analogue multimeter set to 10v, the resistance of the meter will be about 100k, if the meter has a sensitivity of 10k/v and the reading will be incorrect.
Here how it works:
Every meter has a sensitivity. The sensitivity of the meter is the sensitivity of the movement and is the amount of current required to deflect the needle FULL SCALE.
This current is very small, normally 1/10th of a milliamp and corresponds to a sensitivity of 10k/volt (or 1/30th mA, for a sensitivity of 30k/v).
If an analogue meter is set to 10v, the internal resistance of the meter will be 100k for a 10k/v movement.
If this multimeter is used to test the following circuit, the reading will be inaccurate.
The reading should be 5v as show in diagram **A**.
But the analogue multimeter has an internal resistance of 100k and it creates a circuit shown in **C**.
The top 1M and 100k from the meter create a combined PARALLEL resistance of 90k.
This forms a series circuit with the lower 1M and the meter will read less than 1v
If we measure the voltage across the lower 1M, the 100k meter will form a value of resistance with the lower 1M and it will read less than 1v
If the multimeter is 30k/v, the readings will be 2v. See how easy it is to get a totally inaccurate reading.

10v

A — 1M, 5v, 1M, (100k)

B — 10v, 1M, 1M

C — 10v, 100k, 1M, 1M

D — 10v, (100k), 1M, 1M

E — 10v, 1M, 100k, 1M

This introduces two new terms:
HIGH IMPEDANCE CIRCUIT and **"RESISTORS in SERIES and PARALLEL."**

If the reading is taken with a Digital Meter, it will be more accurate as a DMM does not take any current from the circuit (to activate the meter). In other words it has a very HIGH input impedance. Most Digital Multimeters have a fixed input resistance (impedance) of 10M - no matter what scale is selected. That's the reason for choosing a DMM for high impedance circuits. It also gives a reading that is accurate to about 1%.

MEASURING VOLTAGES IN A CIRCUIT

You can take many voltage-measurements in a circuit. You can measure "across" a component, or between any point in a circuit and either the positive rail or earth rail (0v rail). In the following circuit, the 5 most important voltage-measurements are shown. Voltage "A" is across the electret microphone. It should be between 20mV and 500mV. Voltage "B" should be about 0.6v. Voltage "C" should be about half-rail voltage. This allows the transistor to amplify both the positive and negative parts of the waveform. Voltage "D" should be about 1-3v. Voltage "E" should be the battery voltage of 12v.

MEASURING VOLTAGES IN A CIRCUIT

MEASURING CURRENT

You will rarely need to take current measurements, however most multimeters have DC current ranges such as 0.5mA, 50mA, 500mA and 10Amp (via the extra banana socket) and some meters have AC current ranges. Measuring the current of a circuit will tell you a lot of things. If you know the normal current, a high or low current can let you know if the circuit is overloaded or not fully operational.

Current is always measured when the circuit is working (i.e: with power applied).
It is measured IN SERIES with the circuit or component under test.
The easiest way to measure current is to remove the fuse and take a reading across the fuse-holder. Or remove one lead of the battery or turn the project off, and measure across the switch.
If this is not possible, you will need to remove one end of a component and measure with the two probes in the "opening."
Resistors are the easiest things to desolder, but you may have to cut a track in some circuits. You have to get an "opening" so that a current reading can be taken.
The following diagrams show how to connect the probes to take a CURRENT reading. Do not measure the current ACROSS a component as this will create a "short-circuit."
The component is designed to drop a certain voltage and when you place the probes across this component, you are effectively adding a "link" or "jumper" and the voltage at the left-side of the component will appear on the right-side. This voltage may be too high for the circuit being supplied and the result will be damage.

Measuring current through a resistor

Measuring the current of a globe

Do NOT measure the CURRENT of a battery
(by placing the meter directly across the terminals)
A battery will deliver a very HIGH current
and damage the meter

Do not measure the "current a battery will deliver" by placing the probes across the terminals. It will deliver a very high current and damage the meter instantly. There are special battery testing instruments for this purpose.
When measuring across an "opening" or "cut," place the red probe on the wire that supplies the voltage (and current) and the black probe on the other wire. This will produce a "POSITIVE" reading.

A positive reading is an UPSCALE READING and the pointer will move across the scale - to the right. A "NEGATIVE READING" will make the pointer hit the "STOP" at the left of the scale and you will not get a reading. If you are using a Digital Meter, a negative sign "-" will appear on the screen to indicate the probes are around the wrong way. No damage will be caused. It just indicates the probes are connected incorrectly.
If you want an accurate CURRENT MEASUREMENT, use a digital meter.

MEASURING RESISTANCE
Turn a circuit off before measuring resistance.
If any voltage is present, the value of resistance will be incorrect.
In most cases you cannot measure a component while it is in-circuit. This is because the meter is actually measuring a voltage across a component and calling it a "resistance." The voltage comes from the battery inside the meter. If any other voltage is present, the meter will produce a false reading.
If you are measuring the resistance of a component while still "in circuit," (with the power off) the reading will be lower than the true reading.

Measuring resistance

Measuring resistance of a heater
(via the leads)

Measuring the resistance of a piece of resistance-wire

resistance wire

12 V

Power must be disconnected from the circuit

Measuring the resistance of a resistor

Do not measure the "Resistance of a Battery"

1. Do not measure the "resistance of a battery." The resistance of a battery (called the Internal impedance) is not measured as shown in the diagrams above. It is measured by creating a current-flow and measuring the voltage across the battery.

Placing a multimeter set to **resistance** (across a battery) will destroy the meter.
2. Do not try to measure the resistance of any voltage or any "supply."

Resistance is measured in OHMs.
The resistance of a 1cm x 1cm bar, one metre long is 1 ohm.
If the bar is thinner, the resistance is higher. If the bar is longer, the resistance is higher.
If the material of the bar is changed, the resistance is higher.
When carbon is mixed with other elements, its resistance increases and this knowledge is used to make RESISTORS.
Resistors have RESISTANCE and the main purpose of a resistor is to reduce the CURRENT FLOW.
It's a bit like standing on a hose. The flow reduces.
When current flow is reduced, the output voltage is also reduced and that why the water does not spray up so high. Resistors are simple devices but they produce many different effects in a circuit.
A resistor of nearly pure carbon may be 1 ohm, but when non-conducting "impurities" are added, the same-size resistor may be 100 ohms, 1,000 ohms or 1 million ohms.
Circuits use values of less than 1 ohm to more than 22 million ohms.

Resistors are identified on a circuit with numbers and letters to show the exact value of resistance - such as 1k   2k2   4M7
The letter Ω (omega - a Greek symbol) is used to identify the word "Ohm."
but this symbol is not available on some word-processors, so the letter "R" is used.
The letter "E" is also sometimes used and both mean "Ohms."
A one-ohm resistor is written "1R" or "1E." It can also be written "1R0" or "1E0."
A resistor of one-tenth of an ohm is written "0R1" or "0E1."   The letter takes the place of the decimal point.
10 ohms = 10R
100 ohms = 100R
1,000 ohms = 1k (k= kilo = one thousand)
10,000 ohms = 10k
100,000 ohms = 100k
1,000,000 ohms = 1M   (M = MEG = one million)

The size of a resistor has nothing to do with its resistance. The size determines the wattage of the resistor - how much heat it can dissipate without getting too hot.
Every resistor is identified by colour bands on the body, but when the resistor is a surface-mount device, numbers are used and sometimes letters.
You MUST learn the colour code for resistors and the following table shows all the colours for the most common resistors from 1/10th of an ohm to 22 Meg ohms for resistors with 5% and 10% tolerance.

If 3rd band is gold, Divide by 10
If 3rd band is silver, Divide by 100
(to get 0.22ohms etc)

COLOR CODES FOR THE WHOLE E12/E24 RANGE OF RESISTORS
The twelve odd rows - 1, 3, 5...- represent values available in the E12 range only, plus 10M

Reading 4-band resistors

The most "common" type of resistor has 4 bands and is called the 10% resistor. It now has a tolerance of 5% but is still called the "10% type" as the colours increase by 20% so that a resistor can be 10% higher or 10% lower than a particular value and all the resistors produced in a batch can be used.
The first 3 bands produce the resistance and the fourth band is the "tolerance" band.
Gold = 5%
(Silver =10% but no modern resistors are 10%!! - they are 5% 2% or 1%)
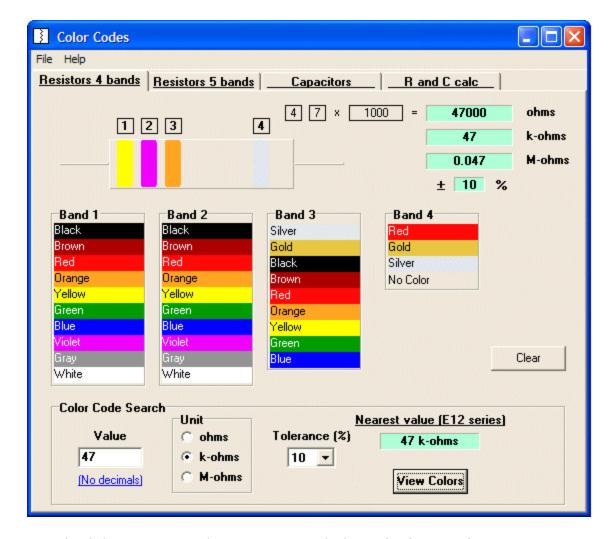
10k	100k	1M0	10M
12k	120k	1M2	22M
15k	150k	1M5	
18k	180k	1M8	
22k	220k	2M2	0R1
27k	270k	2M7	R22
33k	330k	3M3	0R0
39k	390k	3M9	zero ohm (link)
47k	470k	4M7	
56k	560k	5M6	
68k	680k	6M8	
82k	820k	8M2	

4 Bands — 22Ω, 5%

5 Bands — 2200Ω, 1%

6 Bands — 100KΩ, 0.1%, 15ppm

	Silver	Gold	Black 0	Brown 1	Red 2	Orange 3	Yellow 4	Green 5	Blue 6	Purple 7	Grey 8	White 9
Significant Figures			0 0 0	1 1 1	2 2 2	3 3 3	4 4 4	5 5 5	6 6 6	7 7 7	8 8 8	9 9 9
Multiplier	0.01	0.1	1	10	100	1K	10K	100k	1M	10M		
Tolerance	10%	5%		1%	2%			0.5%	0.25%	0.1%		
Temperature Coefficient				100ppm	50ppm	15ppm	25ppm					

Resistor Color Code System

Here is another well-designed resistor colour code chart:

Download the program and save it on your desk-top for future reference:
ColourCode.exe (520KB)
ColourCode.zip   (230KB)
ColourCode.rar   (180KB)

RESISTORS LESS THAN 10 OHMS
When the **third** band is gold, it indicates the value of the "colors" must be divided by 10.
Gold = "divide by 10" to get values 1R0 to 8R2
When the **third** band is silver, it indicates the value of the "colors" must be divided by 100. (Remember: more letters in the word "silver" thus the divisor is "a larger division.")
Silver = "divide by 100" to get values R1 to R82
e.g: 0R1 = 0.1 ohm   0R22 = point 22 ohms
See 4th Column above for examples.

The letters "R, k and M" take the place of a decimal point.
e.g: 1**R**0 = 1 ohm   2**R**2 = 2 point 2 ohms   22**R** = 22 ohms
2**k**2 = 2,200 ohms   100**k** = 100,000 ohms
2**M**2 = 2,200,000 ohms

HOW TO REMEMBER THE COLOUR CODE:
Each colour has a "number" (or divisor) corresponding to it.
Most of the colours are in the same order as in the spectrum. You can see the spectrum in a rainbow. It is: ROY G BIV and the colours for resistors are in the same sequence.
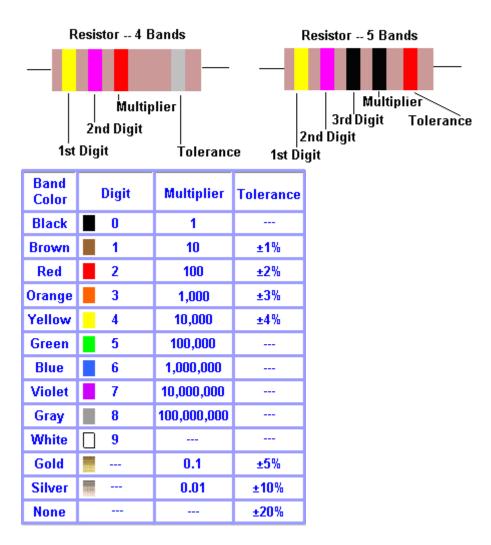
black
brown   - colour of increasing temperature

red
orange
yellow
green
blue
(indigo - that part of the spectrum between blue and violet)
violet
gray
white

colour	value	No of zero's
silver	-2	divide by 100
gold	-1	divide by 10
black	0	No zeros
brown	1	0
red	2	00
orange	3	,000 or k
yellow	4	0,000
green	5	00,000
blue	6	M
violet	7	
gray	8	
white	9	

Band Color	Digit	Multiplier	Tolerance
Black	0	1	---
Brown	1	10	±1%
Red	2	100	±2%
Orange	3	1,000	±3%
Yellow	4	10,000	±4%
Green	5	100,000	---
Blue	6	1,000,000	---
Violet	7	10,000,000	---
Gray	8	100,000,000	---
White	9	---	---
Gold	---	0.1	±5%
Silver	---	0.01	±10%
None	---	---	±20%

Here are some common ways to remember the colour code:

Bad Beer Rots Our Young Guts, But Vodka Goes Well
Bright Boys Rave Over Young Girls But Violet Gets Wed
Bad Boys Rave Over Young Girls But Violet Gets Wed with Gold and Silver.

Reading 5-band resistors:

5-band resistors are easy to read if you remember two simple points. The first three bands provide the digits in the answer and the 4th band supplies the number of zero's.

Reading "STANDARD VALUES" (on 5-band resistors)

5-band resistors are also made in "Standard Values" but will have different colours to 4-band "common" resistors - and will be confusing if you are just starting out. For instance, a 47k 5% resistor with 4-bands will be: yellow-purple-orange-gold. For a 47k 1% resistor the colours will be yellow-purple-black-red-brown. The brown colour-band represents 1%.
The first two colour-bands for a STANDARD VALUE or "common value" in 1% or 5% will be the SAME. These two bands provide the digits in the answer.
It's the 3rd band for a 5% resistor that is expanded into two bands in a 1% resistor. But it's easy to follow.
For a standard value, the 3rd band in a 1% resistor is BLACK. This represents a ZERO in the answer. (For 5-band resistors BLACK represents a ZERO when in the third band. This is different to 4-band resistors where black represents the word OHMS! If the third band is BROWN, the answer will be 1).
So the 4th band has to represent one-less ZERO and is one colour UP THE COLOUR CHART! In other words the 3rd and 4th bands (combined) on a 1% resistor produces the same number of zero's as the 3rd band on a 5% resistor!

Resistors come in a range of values and the two most common are the E12 and E24 series. The E12 series comes in twelve values for each decade. The E24 series comes in twenty-four values per decade.

E12 series - 10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68, 82

E24 series - 10, 11, 12, 13, 15, 16, 18, 20, 22, 24, 27, 30, 33, 36, 39, 43, 47, 51, 56, 62, 68, 75, 82, 91
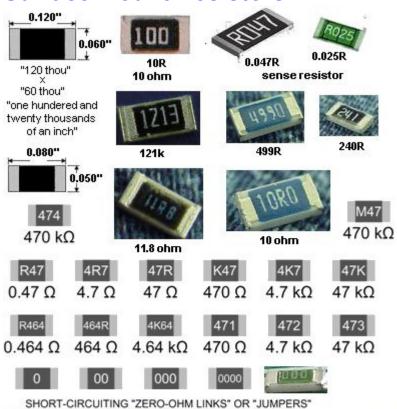
Here is the complete list of 1% 1/4watt resistors from:
CIRCUIT SPECIALISTS. The following list covers 10 ohms (10R) to 1M.
To buy 1% resistors from Circuit Specialists, click: HERE.
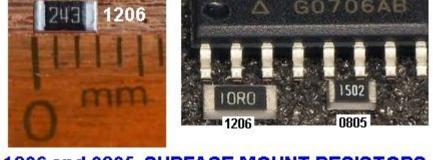
10R	121R	806R	3k83	7k15	14k7	39k2	121k
12R1	150R	825R	3k92	7k5	15k	40k2	147k
15R	182R	909R	4k02	7k87	15k8	44k2	150k
18R2	200R	1k0	4k22	71k5	16k9	46k4	182k
22R1	221R	1k21	4k64	8k06	17k4	47k	200k
27R4	240R	1k5	4k75	8k25	17k8	47k5	212k
30R1	249R	1k82	4k7	8k45	18k2	49k9	221k
33R2	274R	2k	4k87	8k66	20k	51k1	226k
36R5	301R	2k21	4k99	8k87	22k1	53k6	249k
39R2	332R	2k2	5k11	9k09	22k6	56k2	274k
47R5	348R	2k43	5k23	9k31	23k7	61k9	301k
49R9	392R	2k49	5k36	9k53	24k9	68k1	332k
51R1	402R	2k67	5k49	9k76	27k4	69k8	357k
56R2	475R	2k74	5k62	10k	29k4	75k0	392k
68R1	499R	3k01	5k76	11k	30k1	82k5	475k
75R	565R	3k32	5k9	12k	33k2	90k	487k
82R5	604R	3k48	6k04	12k1	34k8	90k9	499k
90R9	681R	3k57	6k19	12k4	36k5	95k3	562k
100R	750R	3k74	6k81	13k	38k3	100k	604k
							1M

Here is the list of 1% resistors from suppliers (such as Farnell):
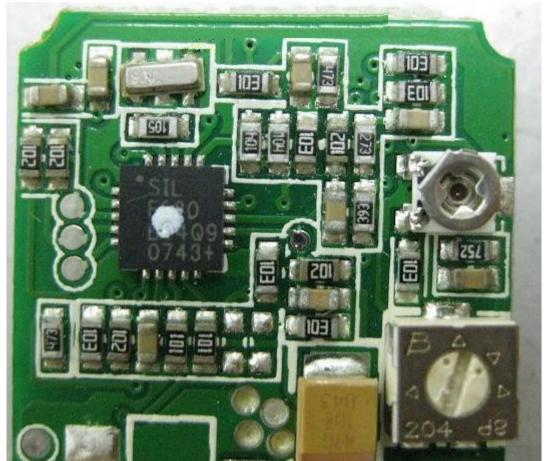
1R0	13R	68R	360R	1k8	9k1	47k	240k
1R2	15R	75R	390R	2k0	10k	51k	270k
1R5	16R	82R	430R	2k2	11k	56k	300k
2R2	18R	91R	470R	2k4	12k	62k	330k
2R7	20R	100R	510R	2k7	13k	68k	360k
3R3	22R	110R	560R	3k	15k	75k	390k
3R9	24R	120R	620R	3k3	16k	82k	430k
4R7	27R	130R	680R	3k6	18k	91k	470k
5R6	30R	150R	750R	3k9	20k	100k	510k
6R2	33R	160R	820R	4k3	22k	110k	560k
6R8	36R	180R	910R	4k7	24k	120k	620k
7R5	39R	200R	1k	5k1	27k	130k	680k
8R2	43R	220R	1k1	5k6	30k	150k	750k
9R1	47R	240R	1k2	6k2	33k	160k	820k
10R	51R	270R	1k3	6k8	36k	180k	910k
11R	56R	300R	1k5	7k5	39k	200k	1M
12R	62R	330R	1k6	8k2	43k	220k	

Surface Mount Resistors
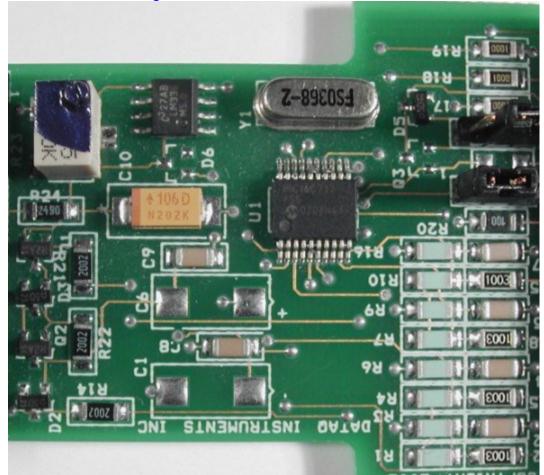
1206 and 0805 SURFACE MOUNT RESISTORS

3-digit Surface Mount resistors on a PC board
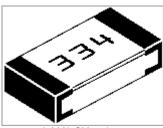
4-digit Surface Mount resistors on a PC board

The photo above shows surface mount resistors on a circuit board. The components

that are not marked are capacitors (capacitors are NEVER marked).
All the SM resistors in the above photos conform to a 3-digit or 4-digit code. But there are a number of codes, and the 4-digit code caters for high tolerance resistors, so it's getting very complicated.
Here is a basic 3-digit SM resistor:

A 330k SM resistor

The first two digits represent the two digits in the answer. The third digit represents the number of zero's you must place after the two digits. The answer will be OHMS. For example: 334 is written 33 0 000. This is written 330,000 ohms. The comma can be replaced by the letter "k". The final answer is: 330k.
222 = 22 00 = 2,200 = 2k2
473 = 47 000 = 47,000 = 47k
474 = 47 0000 = 470,000 = 470k
105 = 10 00000 = 1,000,000 = 1M = one million ohms
There is one trick you have to remember. Resistances less than 100 ohms are written: 100, 220, 470. These are 10 and NO zero's = 10 ohms = 10R
or 22 and no zero's = 22R or 47 and no zero's = 47R. Sometimes the resistor is marked: 10, 22 and 47 to prevent a mistake.

Remember:
R = ohms
k = kilo ohms = 1,000 ohms
M = Meg = 1,000,000 ohms
The 3 letters (R, k and M) are put in place of the decimal point. This way you cannot make a mistake when reading a value of resistance.

Surface Mount CURRENT SENSING Resistors
Many new types of CURRENT SENSING surface-mount resistors are appearing on the market and these are creating lots of new problems.
Fortunately all resistors are marked with the value of resistance and these resistors are identified in MILLIOHMS. A miili ohm is one thousandth or an ohm and is written 0.001 when writing a normal mathematical number.
When written on a surface mount resistor, the letter R indicates the decimal point and it also signifies the word "OHM" or "OHMS" and one milli-ohm is written R001
Five miiliohms is R005   and one hundred milliohms is R100
Some surface mount resistors have the letter "M" after the value to indicate the resistor has a rating of 1 watt. e.g: R100M   These surface-mount resistors are specially-made to withstand a high temperature and a surface-mount resistor of the same size is normally 250mW or less.
These current-sensing resistors can get extremely hot and the PC board can become burnt or damaged.
When designing a PC board, make the lands very large to dissipate the heat.
Normally a current sensing resistor is below one ohm (1R0) and it is easy to identify them as R100 etc.
You cannot measure the value of a current sensing resistor as the leads of a multimeter have a higher resistance than the resistor and few multimeters can read values below one ohm.
If the value is not visible, you will have to refer to the circuit.
Before replacing it, work out why it failed.
Generally it gets too hot. Use a larger size and add tiny heatsinks on each end.
Here are some surface=mount current-sense resistors:

R003
3 milli-ohm

R22
220 milli-ohm

R015
15 milli-ohm

9.1mΩF
9 point 1 milli-ohm

221mΩ F
220 milli-ohm

R005
5 milli-ohm

CSR2
50mΩF
50 milli-ohm 1%

3920C
R002
1 %
2 milli-ohm

R030

DALE
R03F

R030
30 milli-ohm

RO 10
10 milli-ohm

M75
www.bosen.ci
1-watt ????

THE COMPLETE RANGE OF SM RESISTOR MARKINGS

Click to see the complete range of SM resistor markings for 3-digit code:

Click to see the complete range of SM resistor markings for 4-digit code:

0000 is a value on a surface-mount resistor. It is a zero-ohm **LINK**!
Resistances less than 10 ohms have 'R' to indicate the position of the decimal point.
Here are some examples:

Three Digit Examples	Four Digit Examples
330 is 33 ohms - not 330 ohms	1000 is 100 ohms - not 1000 ohms
221 is 220 ohms	4992 is 49 900 ohms, or 49k9
683 is 68 000 ohms, or 68k	1623 is 162 000 ohms, or 162k
105 is 1 000 000 ohms, or 1M	0R56 or R56 is 0.56 ohms
8R2 is 8.2 ohms	

A new coding system has appeared on **1% types**. This is known as the EIA-96 marking method. It consists of a three-character code. The first two digits signify the 3 significant digits of the resistor value, using the lookup table below. The third character - a letter - signifies the multiplier.

code	value	code	value	code	value	code	value	code	value	code	value
01	100	17	147	33	215	49	316	65	464	81	681
02	102	18	150	34	221	50	324	66	475	82	698
03	105	19	154	35	226	51	332	67	487	83	715
04	107	20	158	36	232	52	340	68	499	84	732
05	110	21	162	37	237	53	348	69	511	85	750
06	113	22	165	38	243	54	357	70	523	86	768
07	115	23	169	39	249	55	365	71	536	87	787
08	118	24	174	40	255	56	374	72	549	88	806
09	121	25	178	41	261	57	383	73	562	89	825
10	124	26	182	42	267	58	392	74	576	90	845
11	127	27	187	43	274	59	402	75	590	91	866
12	130	28	191	44	280	60	412	76	604	92	887
13	133	29	196	45	287	61	422	77	619	93	909
14	137	30	200	46	294	62	432	78	634	94	931
15	140	31	205	47	301	63	442	79	649	95	953
16	143	32	210	48	309	64	453	80	665	96	976

The **multiplier** letters are as follows:

letter	mult		letter	mult
F	100000		B	10
E	10000		A	1
D	1000		X or S	0.1
C	100		Y or R	0.01

22A is a 165 ohm resistor, **68C** is a 49900 ohm (49k9) and **43E** a 2740000 (2M74). This marking scheme applies to 1% resistors only.

A similar arrangement can be used for **2% and 5%** tolerance types. The multiplier letters are identical to 1% ones, but occur **before** the number code and the following **code** is used:

2%				5%			
code	value	code	value	code	value	code	value
01	100	13	330	25	100	37	330
02	110	14	360	26	110	38	360
03	120	15	390	27	120	39	390
04	130	16	430	28	130	40	430
05	150	17	470	29	150	41	470
06	160	18	510	30	160	42	510
07	180	19	560	31	180	43	560

08	200	20	620	32	200	44	620
09	220	21	680	33	220	45	680
10	240	22	750	34	240	46	750
11	270	23	820	35	270	47	820
12	300	24	910	36	300	48	910

With this arrangement, **C31** is 5%, 18000 ohm (18k), and **D18** is 510000 ohms (510k) 2% tolerance.

Always check with an ohm-meter (a multimeter) to make sure.

Chip resistors come in the following styles and ratings:
Style: 0402, 0603, 0805, 1206, 1210, 2010, 2512, 3616, 4022
Power Rating: 0402(1/16W), 0603(1/10W), 0805(1/8W), 1206(1/4W), 1210(1/3W), 2010(3/4W), 2512(1W), 3616(2W), 4022(3W)
Tolerance: 0.1%, 0.5%, 1%, 5%
Temperature Coefficient: 25ppm 50ppm 100ppm

EIA marking code for surface mount (SMD) resistors

01S = 1R	01R = 10R	01A = 100R	01B = 1k	01C = 10k	01D = 100k	01E = 1M	01F = 10M
02S = 1R02	02R = 10R2	02A = 102R	02B = 1k02	02C = 10k2	02D = 102k	02E = 1M02	
03S = 1R05	03R = 10R5	03A = 105R	03B = 1k05	03C = 10k5	03D = 105k	03E = 1M05	18F = 15M
04S = 1R07	04R = 10R7	04A = 107R	04B = 1k07	04C = 10k7	04D = 107k	04E = 1M07	
05S = 1R1	05R = 11R	05A = 110R	05B = 1k1	05C = 11k	05D = 110k	05E = 1M1	30F = 20M
06S = 1R13	06R = 11R3	06A = 113R	06B = 1k13	06C = 11k3	06D = 113k	06E = 1M13	
07S = 1R15	07R = 11R5	07A = 115R	07B = 1k15	07C = 11k5	07D = 115k	07E = 1M15	
08S = 1R18	08R = 11R8	08A = 118R	08B = 1k18	08C = 11k8	08D = 118k	08E = 1M18	
09S = 1R21	09R = 12R1	09A = 121R	09B = 1k21	09C = 12k1	09D = 121k	09E = 1M21	
10S = 1R24	10R = 12R4	10A = 124R	10B = 1k24	10C = 12k4	10D = 124k	10E = 1M24	
11S = 1R27	11R = 12R7	11A = 127R	11B = 1k27	11C = 12k7	11D = 127k	11E = 1M27	
12S = 1R3	12R = 13R	12A = 130R	12B = 1k3	12C = 13k	12D = 130k	12E = 1M3	
13S = 1R33	13R = 13R3	13A = 133R	13B = 1k33	13C = 13k3	13D = 133k	13E = 1M33	
14S = 1R37	14R = 13R7	14A = 137R	14B = 1k37	14C = 13k7	14D = 137k	14E = 1M37	
15S = 1R4	15R = 14R	15A = 140R	15B = 1k4	15C = 14k	15D = 140k	15E = 1M4	
16S = 1R43	16R = 14R3	16A = 143R	16B = 1k43	16C = 14k3	16D = 143k	16E = 1M43	
17S = 1R47	17R = 14R7	17A = 147R	17B = 1k47	17C = 14k7	17D = 147k	17E = 1M47	
18S = 1R5	18R = 15R	18A = 150R	18B = 1k5	18C = 15k	18D = 15k	18E = 1M5	
19S = 1R54	19R = 15R4	19A = 154R	19B = 1k54	19C = 15k4	19D = 154k	19E = 1M54	
20S = 1R58	20R = 15R8	20A = 158R	20B = 1k58	20C = 15k8	20D = 158k	20E = 1M58	
21S = 1R62	21R = 16R2	21A = 162R	21B = 1k62	21C = 16k2	21D = 162k	21E = 1M62	
22S = 1R65	22R = 16R5	22A = 165R	22B = 1k65	22C = 16k5	22D = 165k	22E = 1M65	
23S = 1R69	23R = 16R9	23A = 169R	23B = 1k69	23C = 16k9	23D = 169k	23E = 1M69	
24S = 1R74	24R = 17R4	24A = 174R	24B = 1k74	24C = 17k4	24D = 174k	24E = 1M74	
25S = 1R78	25R = 17R8	25A = 178R	25B = 1k78	25C = 17k8	25D = 178k	25E = 1M78	
26S = 1R82	26R = 18R2	26A = 182R	26B = 1k82	26C = 18k2	26D = 182k	26E = 1M82	
27S = 1R87	27R = 18R7	27A = 187R	27B = 1k87	27C = 18k7	27D = 187k	27E = 1M87	
28S = 1R91	28R = 19R1	28A = 191R	28B = 1k91	28C = 19k1	28D = 191k	28E = 1M91	
29S = 1R96	29R = 19R6	29A = 196R	29B = 1k96	29C = 19k6	29D = 196k	29E = 1M96	
30S = 2R0	30R = 20R0	30A = 200R	30B = 2k0	30C = 20k0	30D = 200k	30E = 2M0	
31S = 2R05	31R = 20R5	31A = 205R	31B = 2k05	31C = 20k5	31D = 205k	31E = 2M05	
32S = 2R10	32R = 21R0	32A = 210R	32B = 2k10	32C = 21k0	32D = 210k	32E = 2M10	
33S = 2R15	33R = 21R5	33A = 215R	33B = 2k15	33C = 21k5	33D = 215k	33E = 2M15	
34S = 2R21	34R = 22R1	34A = 221R	34B = 2k21	34C = 22k1	34D = 221k	34E = 2M21	
35S = 2R26	35R = 22R6	35A = 226R	35B = 2k26	35C = 22k6	35D = 226k	35E = 2M26	
36S = 2R32	36R = 23R2	36A = 232R	36B = 2k32	36C = 23k2	36D = 232k	36E = 2M32	
37S = 2R37	37R = 23R7	37A = 237R	37B = 2k37	37C = 23k7	37D = 237k	37E = 2M37	
38S = 2R43	38R = 24R3	38A = 243R	38B = 2k43	38C = 24k3	38D = 243k	38E = 2M43	
39S = 2R49	39R = 24R9	39A = 249R	39B = 2k49	39C = 24k9	39D = 249k	39E = 2M49	
40S = 2R55	40R = 25R5	40A = 255R	40B = 2k55	40C = 25k5	40D = 255k	40E = 2M55	
41S = 2R61	41R = 26R1	41A = 261R	41B = 2k61	41C = 26k1	41D = 261k	41E = 2M61	
42S = 2R67	42R = 26R7	42A = 267R	42B = 2k67	42C = 26k7	42D = 267k	42E = 2M67	
43S = 2R74	43R = 27R4	43A = 274R	43B = 2k74	43C = 27k4	43D = 274k	43E = 2M74	
44S = 2R80	44R = 28R0	44A = 280R	44B = 2k80	44C = 28k0	44D = 280k	44E = 2M80	
45S = 2R87	45R = 28R7	45A = 287R	45B = 2k87	45C = 28k7	45D = 287k	45E = 2M87	
46S = 2R94	46R = 29R4	46A = 294R	46B = 2k94	46C = 29k4	46D = 294k	46E = 2M94	
47S = 3R01	47R = 30R1	47A = 301R	47B = 3k01	47C = 30k1	47D = 301k	47E = 3M01	
48S = 3R09	48R = 30R9	48A = 309R	48B = 3k09	48C = 30k9	48D = 309k	48E = 3M09	
49S = 3R16	49R = 31R6	49A = 316R	49B = 3k16	49C = 31k6	49D = 316k	49E = 3M16	
50S = 3R24	50R = 32R4	50A = 324R	50B = 3k24	50C = 32k4	50D = 324k	50E = 3M24	
51S = 3R32	51R = 33R2	51A = 332R	51B = 3k32	51C = 33k2	51D = 332k	51E = 3M32	
52S = 3R4	52R = 34R0	52A = 340R	52B = 3k4	52C = 34k0	52D = 340k	52E = 3M4	
53S = 3R48	53R = 34R8	53A = 348R	53B = 3k48	53C = 34k8	53D = 348k	53E = 3M48	

54S = 3R57	54R = 35R7	54A = 357R	54B = 3k57	54C = 35k7	54D = 357k	54E = 3M57
55S = 3R65	55R = 36R5	55A = 365R	55B = 3k65	55C = 36k5	55D = 365k	55E = 3M65
56S = 3R74	56R = 37R4	56A = 374R	56B = 3k74	56C = 37k4	56D = 374k	56E = 3M74
57S = 3R83	57R = 38R3	57A = 383R	57B = 3k83	57C = 38k3	57D = 383k	57E = 3M83
58S = 3R92	58R = 39R2	58A = 392R	58B = 3k92	58C = 39k2	58D = 392k	58E = 3M92
59S = 4R02	59R = 40R2	59A = 402R	59B = 4k02	59C = 40k2	59D = 402k	59E = 4M02
60S = 4R12	60R = 41R2	60A = 412R	60B = 4k12	60C = 41k2	60D = 412k	60E = 4M12
61S = 4R22	61R = 42R2	61A = 422R	61B = 4k22	61C = 42k2	61D = 422k	61E = 4M22
62S = 4R32	62R = 43R2	62A = 432R	62B = 4k32	62C = 43k2	62D = 432k	62E = 4M32
63S = 4R42	63R = 44R2	63A = 442R	63B = 4k42	63C = 44k2	63D = 442k	63E = 4M42
64S = 4R53	64R = 45R3	64A = 453R	64B = 4k53	64C = 45k3	64D = 453k	64E = 4M53
65S = 4R64	65R = 46R4	65A = 464R	65B = 4k64	65C = 46k4	65D = 464k	65E = 4M64
66S = 4R75	66R = 47R5	66A = 475R	66B = 4k75	66C = 47k5	66D = 475k	66E = 4M75
67S = 4R87	67R = 48R7	67A = 487R	67B = 4k87	67C = 48k7	67D = 487k	67E = 4M87
68S = 4R99	68R = 49R9	68A = 499R	68B = 4k99	68C = 49k9	68D = 499k	68E = 4M99
69S = 5R11	69R = 51R1	69A = 511R	69B = 5k11	69C = 51k1	69D = 511k	69E = 5M11
70S = 5R23	70R = 52R3	70A = 523R	70B = 5k23	70C = 52k3	70D = 523k	70E = 5M23
71S = 5R36	71R = 53R6	71A = 536R	71B = 5k36	71C = 53k6	71D = 536k	71E = 5M36
72S = 5R49	72R = 54R9	72A = 549R	72B = 5k49	72C = 54k9	72D = 549k	72E = 5M49
73S = 5R62	73R = 56R2	73A = 562R	73B = 5k62	73C = 56k2	73D = 562k	73E = 5M62
74S = 5R76	74R = 57R6	74A = 576R	74B = 5k76	74C = 57k6	74D = 576k	74E = 5M76
75S = 5R9	75R = 59R0	75A = 590R	75B = 5k9	75C = 59k0	75D = 590k	75E = 5M9
76S = 6R04	76R = 60R4	76A = 604R	76B = 6k04	76C = 60k4	76D = 604k	76E = 6M04
77S = 6R19	77R = 61R9	77A = 619R	77B = 6k19	77C = 61k9	77D = 619k	77E = 6M19
78S = 6R34	78R = 63R4	78A = 634R	78B = 6k34	78C = 63k4	78D = 634k	78E = 6M34
79S = 6R49	79R = 64R9	79A = 649R	79B = 6k49	79C = 64k9	79D = 649k	79E = 6M49
80S = 6R65	80R = 66R5	80A = 665R	80B = 6k65	80C = 66k5	80D = 665k	80E = 6M65
81S = 6R81	81R = 68R1	81A = 681R	81B = 6k81	81C = 68k1	81D = 681k	81E = 6M81
82S = 6R98	82R = 69R8	82A = 698R	82B = 6k98	82C = 69k8	82D = 698k	82E = 6M98
83S = 7R15	83R = 71R5	83A = 715R	83B = 7k15	83C = 71k5	83D = 715k	83E = 7M15
84S = 7R32	84R = 73R2	84A = 732R	84B = 7k32	84C = 73k2	84D = 732k	84E = 7M32
85S = 7R5	85R = 75R0	85A = 750R	85B = 7k5	85C = 75k0	85D = 750k	85E = 7M5
86S = 7R68	86R = 76R8	86A = 768R	86B = 7k68	86C = 76k8	86D = 768k	86E = 7M68
87S = 7R87	87R = 78R7	87A = 787R	87B = 7k87	87C = 78k7	87D = 787k	87E = 7M87
88S = 8R06	88R = 80R6	88A = 806R	88B = 8k06	88C = 80k6	88D = 806k	88E = 8M06
89S = 8R25	89R = 82R5	89A = 825R	89B = 8k25	89C = 82k5	89D = 825k	89E = 8M25
90S = 8R45	90R = 84R5	90A = 845R	90B = 8k45	90C = 84k5	90D = 845k	90E = 8M45
91S = 8R66	91R = 86R6	91A = 866R	91B = 8k66	91C = 84k6	91D = 866k	91E = 8M66
92S = 8R87	92R = 88R7	92A = 887R	92B = 8k87	92C = 88k7	92D = 887k	92E = 8M87
93S = 9R09	93R = 90R9	93A = 909R	93B = 9k09	93C = 90k9	93D = 909k	93E = 9M09
94S = 9R31	94R = 93R1	94A = 931R	94B = 9k31	94C = 93k1	94D = 931k	94E = 9M31
95S = 9R53	95R = 95R3	95A = 953R	95B = 9k53	95C = 95k3	95D = 953k	95E = 9M53
96S = 9R76	96R = 97R6	96A = 976R	96B = 9k76	96C = 97k6	96D = 976k	96E = 9M76

 If you want an accurate RESISTANCE measurement, remove the resistor from the circuit and use a Digital meter.

SURFACE MOUNT COMPONENTS - PACKS

Talking Electronics has packs of components for the repairman. The following packs are available:

SURFACE MOUNT RESISTOR PACK consists of 1 off each standard value
10 ohms to 1M & 2M2 (60 resistors)
$14.20   including pack and post

SURFACE MOUNT CAPACITOR PACK consists of:
2 - 10p   5 - 47p   5 - 100p   5 - 470p   5 - 1n   5 - 10n   5 - 22n   5 - 100n
5 - 1u 16v electrolytic   5 - 10u 16v electrolytic
(40 components)
$23.80 including pack and post

SURFACE MOUNT DIODE PACK consists of:
5 - 1N 4148 (marked as "A6")
$10.00   including pack and post

SURFACE MOUNT TRANSISTOR PACK consists of:
5 - BC 848 (marked as "1K") NPN
5 - BC858 PNP
$10.00   including pack and post

CREATING ANY VALUE OF RESISTANCE

Any value of resistance can be created by connecting two resistors in PARALLEL or SERIES.
You can also create a higher wattage resistor by connecting them in SERIES OR PARALLEL.
We are only going to cover two EQUAL VALUE resistors in SERIES or in PARALLEL.
If you want to create a "Special Value," simply connect two resistors and read the value with a Digital Meter. Keep changing the values until you get the required value.
We are not going into series or Parallel formulae. You can easily find a value with a multimeter.
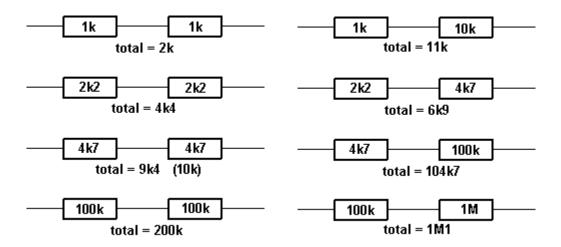
TWO EQUAL-VALUE RESISTORS IN SERIES
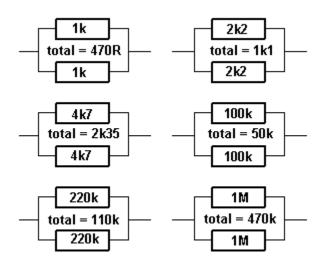
Two equal-value resistors IN SERIES creates a value of DOUBLE. You simply ADD the values.
This can be done with any to two values as shown. Three equal-value resistors in series is three times the value.

1k — 1k	1k — 10k
total = 2k	total = 11k
2k2 — 2k2	2k2 — 4k7
total = 4k4	total = 6k9
4k7 — 4k7	4k7 — 100k
total = 9k4 (10k)	total = 104k7
100k — 100k	100k — 1M
total = 200k	total = 1M1

TWO EQUAL-VALUE RESISTORS IN PARALLEL

Two equal-value resistors IN PARALLEL creates a value of HALF. Three equal-value resistors in parallel is equal to one-third the value.

1k / 1k	2k2 / 2k2
total = 470R	total = 1k1
4k7 / 4k7	100k / 100k
total = 2k35	total = 50k
220k / 220k	1M / 1M
total = 110k	total = 470k

If you want a particular value and it is not available, here is a chart.
Use 2 resistors in series or parallel as shown:

Required Value	R1	Series/ Parallel	R2	Actual value:

10	4R7	S	4R7	9R4
12	10	S	2R2	12R2
15	22	P	47	14R9
18	22	P	100	18R
22	10	S	12	22
27	22	S	4R7	26R7
33	22	S	10	32R
39	220	P	47	38R7
47	22	S	27	49
56	47	S	10	57
68	33	S	33	66
82	27	S	56	83

There are other ways to combine 2 resistors in parallel or series to get a particular value. The examples above are just one way. 4R7 = 4.7 ohms

TESTING A RESISTOR
To check the value of a resistor, it should be removed from the circuit. The surrounding components can affect the reading and make it lower.
Resistors **VERY RARELY** change value, but if it is overheated or damaged, the resistance can increase. You can take the reading of a resistor "in-circuit" in one direction then the other, as the surrounding components may have diodes and this will alter the reading.
You can also test a resistor by feeling its temperature-rise. It is getting too hot if you cannot hold your finger on it (some "metal film" resistors are designed to tolerate quite high temperatures).

TESTING AN "AC" RESISTOR
There is no such thing as an "AC" resistor. Resistors are just "resistors" and they can be in AC circuits or DC circuits. Resistors can be given names such as "Safety Resistor" "Ballast Resistor" "LOAD Resistor" "Feed Resistor"   "Dropper Resistor" or "Supply Resistor." These are just normal resistors with a normal resistance - except a "Safety Resistor."
A safety resistor is made of a flame-proof material such as metal-oxide-film and not carbon-composition. It is designed to "burn out" when too much current flows BUT NOT CATCH FIRE.
It is a low-value resistor and has a voltage-drop across it but this is not intentional. The voltage-drop is to create a "heating-effect" to burn out the resistor. In all the other types of resistor, the voltage-drop is intentional.
A Ballast resistor is a normal resistor and can be called a Power resistor, Dropper resistor, Supply resistor or Feed resistor. It is designed to reduce the voltage from one source and deliver a lower voltage. It is a form of: "in-line" resistor.
A Load Resistor is generally connected across the output of a circuit and turns the energy it receives, into heat.
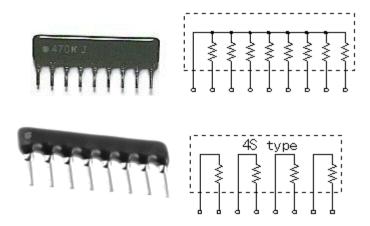
RESISTOR NETWORKS
To reduce the number of resistors in a circuit, some engineers use a set of identical resistors in a package called a Single-In-Line (SIL) resistor network. It is made with many resistors of the same value, all in one package. One end of each resistor is connected all the other resistors and this is the common pin, identified as pin 1 and has a dot on the package.
These packages are very reliable but to make sure all the resistors are as stated, you need to locate pin 1. All values will be identical when referenced to this pin.

RESISTOR NETWORKS

Some resistor networks have a "4S" printed on the component. The 4S indicates the package contains 4 independent resistors that are not wired together inside. The housing has eight leads as shown in the second image.
Independent resistors have an even number of pins and measuring between each pair will produce identical values. Resistance between any pair will indicate leakage and may be a fault.

WIRE WOUND RESISTOR

A wire wound resistor is also called a POWER RESISTOR. This type of resistor can have a resistance as low as 0.1 ohms (one-tenth of an ohm) or as high as about 10k.
The image shows a 0.68 ohm resistor as the letter "R" represents the DECIMAL POINT and R68 is the same a .68 and this is 0.68 ohms. The wattage is 9 watts.
This resistor will allow xxx amps to flow.   To work out the current, use the formula:
Power = Current x Current x resistance
   9   = Current x Current x .68
Divide both sides by 0.68
13.2   = Current x Current
Find the square root of 13.2
Current = 3.6 amps

When 3.6 amps flow through the resistor, the voltage appearing across it will be:
V = current x resistance
   = 3.6   x 0.68
   = 2.5v and the wattage (heat) loss will be 9 watts.

The purpose of a resistor like this is to stop or reduce "ripple." Ripple is the noise or hum in an amplifier when the sound is turned up.
There are many reasons why you need to reduce the level of hum and this resistor will remove ripple as large as 2.5v when 3.6 amps is flowing, provided you have filter electrolytics on both side of the resistor to assist in removing the ripple.

If the letter "R" is in a different position, the value of resistance would be:

68R = 68Ω
6R8 = 6.8Ω
R68 = 0.68Ω

If you replace the R68 resistor a 6R8 resistor by mistake, the voltage across it will rise to 25v and if 3.6 amps flows, the wattage will be: 90 watts!!!
The resistor will glow red and burn out.

TESTING A POSISTOR
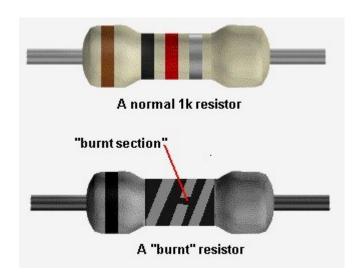
A Posistor is a resistor that connects in series with the degaussing coil around the picture tube or **Monitor.** When cold, it has a very low resistance and a large current flows when the monitor or TV is switched on. This current heats up the Posistor and the resistance increases. This causes the current to decrease and any magnetism in the shadow mask is removed. The posistor can one or two elements and it is kept warm so the resistance remains high. Many Posistors have a second element inside the case that connects directly to the supply to keep the Positive Temperature Coefficient resistor high so that the current through the degaussing coil falls to almost zero. This constant heat eventually destroys the package.
The heavy current that flows when a set is turned ON also causes the posistor to crack and break and this results in poor purity on the screen - as the shadow mask gradually becomes magnetic..
Posistors have different resistance values from different manufacturers and must be replaced with an identical type.
They can be checked for very low resistance when cold but any loose pieces inside the case will indicate a damaged component.

A "BURNT" RESISTOR - normally and technically called a "burnt-out" resistor.

The resistance of a "burnt" resistor can sometimes be determined by scraping away the outer coating - if the resistor has a spiral of resistance-material. You may be able to find a spot where the spiral has been damaged.

A normal 1k resistor

"burnt section"

A "burnt" resistor

Note the spirals of conductive carbon.
The number of spirals has nothing to with the resistance.
It is the amount of carbon particles in the "track" that
determines the resistance. It is also the thickness and width
of the track that determines the resistance.
And then it is the overall size of the resistor that determines the wattage.
And then the size of the leads, the closeness to the PCB and
the size of the lands that eventually determines how hot the resistor
will get.

Clean the "spot" (burnt section of the spiral) very carefully and make sure you can get a good contact with the spiral and the tip of your probe. Measure from one lead of the resistor to the end of the damaged spiral. Then measure from the other lead to the other end of the spiral.
Add the two values and you have an approximate value for the resistor. You can add a small amount for the damaged section.
This process works very well for damaged wire-wound resistors. They can be pulled apart and each section of the resistance-wire (nichrome wire) measured and added to get the full resistance.

There is another way to determine the value of a damaged resistor.
Get a set of resistors of the same wattage as the damaged component and start with a high value. It's handy to know if the resistor is in the range: 10ohm to 100ohms or 1k to 10k etc, but this is not essential.
Start with a very high value and turn the circuit ON. You can perform voltage tests and if you know the expected output voltage, decrease the resistance until this voltage is obtained.
If you do not know the expected voltage, keep reducing the value of resistance until the circuit works as designed.
This is the best advice in a situation where you do not know the value of a resistor.

There is a third way to determine the value and this requires measuring the voltage drop across the resistor and the current-flow. By multiplying the two you will get a wattage and this must be less than the wattage of the resistor being replaced.

TESTING POTENTIOMETERS (variable resistors)
To check the value of a variable resistor, it should be removed from circuit or at least 2 legs should be removed. A Rheostat is a variable resistor using only one end and the middle connected to a circuit.
The resistance between the two outside pins is the value marked on the component and the centre leg will change from nearly zero to the full resistance as the shaft is rotated.
"Pots" generally suffer from "crackle" when turned and this can be fixed by spraying up the shaft and into the pot via the shaft with a tube fixed to a can of "spray-lubricant" (contact cleaner).
"Pre-set pots" and "trim pots" are miniature versions of a potentiometer and they are all tested the same. The photo shows a pot, two mini pots and 3 mini trim pots.
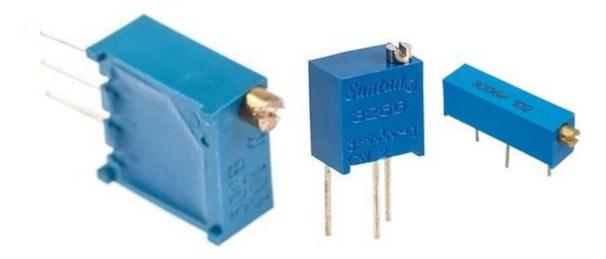
10-Turn POTs

A 10-turn pot is one of the worst items to be designed. I remove them immediately from any design.

You don't know the position of the wiper. You don't know which way you are turning the wiper and you can't remember which way you turned the post "the last time."

The screwdriver always falls out of the slot.

If you need fine adjustment, place fixed resistors on each side of the pot and use a normal mini trim pot with much less resistance.

FOCUS POTS

Focus pots quite often get a spot of dirt where the wiper touches the track. Cleaning with spray fixes the bad focus but if the pot is leaking to chassis from inside the pot (due to the high voltage on the terminals) simply remove it from the chassis and leave it floating (this will restore the high voltage to the picture tube) or you can use one from an old chassis.

MAKING YOUR OWN RESISTOR, CAPACITOR, INDUCTOR or DIODE

Quite often you will not have the exact value of resistance or capacitance for a repair.
We have already covered placing resistors and capacitors in parallel and series:

Resistors in Parallel and/or Series
Capacitors in Parallel and/or Series

Here are some extras:

RESISTORS
Two 1k 0.5watt resistors in parallel produces a 470R 1watt resistor.
Two 1k 0.5watt resistors in series produces a 2k 1watt resistor.

CAPACITORS
Two 100n 100v capacitors in series produces a 50n capacitor @200v

INDUCTORS: Two inductors in series - **ADD THE VALUES**

DIODES: Two 1Amp 400v diodes in series produces a 1Amp 800v diode
              Two 1Amp 400v diodes in parallel produces a 2Amp 400v diode

ZENER DIODES: Zener diodes can be connected in series to get a higher voltage.
        Two 12v zener diodes in series produces a 24v zener.

CONTINUITY
Some multimeters have a "buzzer" that detects when the probes are touching each other or the resistance between the probes is very LOW. This is called a CONTINUITY TESTER.
You can use the resistance scale "x1" or "x10" to detect low values of resistance.
Set the pointer to "0" (right end of the scale) by touching the probes together and adjusting the "zero ohms" control.
When taking a reading, you will have to decide if a low value of resistance is a short-circuit or an "operating value."
For instance, the cold resistance of a 12v car globe is very low (about 2 ohms) and it increases (about 6 times) to 12 ohms when hot.
The "resistance of a circuit" may be very low as the electrolytics in the circuit are uncharged. This may not indicate a true "short-circuit."
The measurement across a diode is not a resistance-value but a "voltage-drop" and that is why the needle swings nearly full-scale.
Leads and wires and cords have a small resistance and depending on the length of the lead, this small resistance may be affecting a circuit.
Remember this:
When a circuit takes 1 amp, and the resistance of the leads is 1 ohm, the voltage drop across the leads will be 1v.
That's why a 12v battery supplying a circuit with these leads will have 11v at the circuit.
Note:
Turn off the equipment before making any continuity tests. The presence of even a small voltage (from an electrolytic) can give a false reading.
You can determine the resistance of a lead very accurately by taking the example above and applying it to your circuit.
If the battery is 12.6v and the voltage across the circuit is 10v, when the current is 2.6 amps, the resistance of the "leads" is 12.6 - 10 = 2.6   R=V/I = 2.6/2.6 = 1ohm. By making the lead shorter or using thicker wire, the resistance will be less and the voltage on the project will increase.
When taking readings in a circuit that has a number of diodes built-into IC's (Integrated Circuits) and transistors, some Continuity Testers will beep and give a false reading.
The following circuit has the advantage of providing a beep when a short-circuit is detected but does not detect the small voltage drop across a diode. This is ideal when testing logic circuits as it is quick and you can listen for the beep while
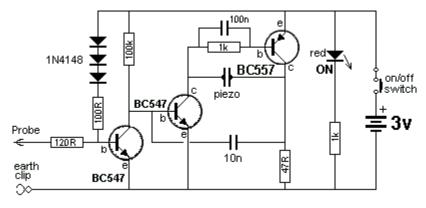
concentrating on the probe. Using a multimeter is much slower.

CONTINUITY TESTER

You can build the circuit on Matrix Board and add it to your Test Equipment.
You will need lots of "Test Equipment" and they can be built from circuits in this eBook.

TESTING FUSES, LEADS AND WIRES

All these components come under the heading TESTING for CONTINUITY. Turn off all power to the equipment before testing for shorts and continuity. Use the low resistance "Ohms Scale" or CONTINUITY range on your multimeter. All fuses, leads and wires should have a low, very low or zero resistance. This proves they are working.

A BLOWN FUSE
The appearance of a fuse after it has "blown" can tell you a lot about the fault in the circuit.
If the inside of the glass tube (of the fuse) is totally blackened, the fuse has been damaged very quickly. This indicates a very high current has passed through the fuse.
Depending on the rating of the fuse, (current rating) you will be able to look for components that can pass a high current when damaged - such as high power transistors, FETs, coils, electrolytics. Before re-connecting the supply, you should test the "SUPPLY RAILS" for resistance. This is done by measuring them on a low OHMs range in one direction then reverse the leads to see if the resistance is low in the other direction.
A reading can be very low at the start because electrolytics need time to charge-up and if the reading gradually increases, the power rail does not have a short. An overload can occur when the supply voltage rises to nearly full voltage, so you sometimes have to fit a fuse and see how long it takes to "blow."
If the fuse is just slightly damaged, you will need to read the next part of this eBook, to see how and why this happens:

FAST AND SLOW BLOW FUSES
There are many different sizes, shapes and ratings of a fuse. They are all current ratings as a fuse does not have a voltage rating. Some fuses are designed for cars as they fit into the special fuse holders. A fuse can be designed for 50mA, 100mA, 250mA, 315mA, 500mA, 1Amp, 1.5amp, 2amp, 3amp, 3.15amp 5amp, 10amp, 15amp, 20amp, 25amp, 30amp, 35amp, 50amp and higher.
Some fuses are fast-blow and some are slow-blow.
A "normal" fuse consists of a length of thin wire. Or it may be a loop of wire that is thin near the middle of the fuse. This is the section that will "burn-out."
A "normal" fuse is a fast-blow fuse. For instance, a 1amp fuse will remain intact when up to 1.25 amp flows. When a circuit is turned on, it may take 2-3 amps for a very short period of time and a normal 1 amp fuse will get very hot and the wire will stretch but not "burn-out." You can see the wire move when the supply turns on.
If the current increases to 2amps, the fuse will still remain intact. It needs about 3 amp to heat up the wire to red-hot and burn out.

If the current increases to 5 amp, the wire VOLATILISES (burns-out) and deposits carbon-black on the inside of the glass tube.
A slow-blow fuse uses a slightly thicker piece of wire and the fuse is made of two pieces of wire joined in the middle with a dob of low-temperature solder. Sometimes one of the pieces of wire is a spring and when the current rises to 2.5 amp, the heat generated in the wire melts the solder and the two pieces of wire "spring apart."
A slow-blow fuse will allow a higher current-surge to pass through the fuse and the wire will not heat up and sag.
Thus the fuse is not gradually being damaged and it will remain in a perfect state for a long period of time.
A fuse does not protect electronic equipment from failing. It acts AFTER the equipment has failed.
It will then protect a power supply from delivering a high current to a circuit that has failed.
If a slow-blow fuse has melted the solder, it could be due to a slight overload, slight weakening of the fuse over a period of time or the current-rating may be too low.
You can try another fuse to see what happens.
You can replace a fast-acting fuse (normal fuse) with a slow blow if the fast-acting fuse has been replaced a few times due to deterioration when the equipment is turned on.
But you cannot replace a slow-blow fuse with a fast acting fuse as it will be damaged slightly each time the equipment is turned on and eventually fail.

100mA FUSES

Fuses below about 100mA are very hard to make and very unreliable.
Many circuits take a high current when turned to charge the electrolytics and a 100mA (or 50mA or 63mA fuse) will bow and stretch and change shape, every time the equipment is turned ON.
Eventually it will break, due to it heating-up and stretching.
To produce a reliable fuse below 100mA, some manufacturers have placed a resistor inside the fuse and connected it to a spring. One end of the resistor is soldered to a wire with low-temperature metal and when the resistor gets hot, the metal softens and the spring pulls the resistor away from the wire.
Quite often you can heat up the metal and connect the wire and the fuse is perfect.
This type of fuse is called a DELAY fuse and the current rating is shown on the end-cap.
The value of the resistor determines the current rating.
There is a small voltage across this type of fuse and it means the circuit sees a slightly lower voltage than the supply voltage.
The third photo shows the pot of solder or low-temp metal and a wire connected to a spring. The heat generated in the wire is passed to the solder and it softens. The spring pulls the two components apart. You can smash the glass and set up the fuse in the two fuse-holders and repair the fuse while you wait for a new fuse to arrive.

TESTING COILS, INDUCTORS and YOKES

Coils, inductors, chokes and yokes are just coils (turns) of wire. The wire may be wrapped around a core made of iron or ferrite.
It is labeled "L" on a circuit board.
You can test this component for continuity between the ends of the winding and also make sure there is no continuity between the winding and the core.
The winding can be less than one ohm, or greater than 100 ohms. A coil of wire is also called an INDUCTOR and it might look like a very simple component, but it can operate in a very complex way.
The way it works is a discussion for another eBook. It is important to understand the turns are insulated but a slight fracture in the insulation can cause two turns to touch each other and this is called a "SHORTED TURN" or you can say the inductor has "SHORTED TURNS."
When this happens, the inductor allows the circuit to draw MORE CURRENT. This causes the fuse to "blow."
The quickest way to check an inductor is to replace it, but if you want to measure the inductance, you can use an INDUCTANCE METER. You can then compare the inductance with a known good component.
An inductor with a shorted turn will have a very low or zero inductance, however you may not be able to detect the fault when it is not working in a circuit as the fault may be created by a high voltage generated between two of the turns.
Faulty yokes (both horizontal and vertical windings) can cause the picture to reduce in size and/or bend or produce a single horizontal line.
A TV or monitor screen is the best piece of Test Equipment as it has identified the fault. It is pointless trying to test the windings further as you will not be able to test them under full operating conditions. The fault may not show up when a low voltage (test voltage) is applied.

MEASURING AND TESTING INDUCTORS

Inductors are measured with an INDUCTANCE METER but the value of some inductors is very small and some Inductance Meters do not give an accurate reading. The solution is to measure a larger inductor and note the reading. Now put the two inductors in SERIES and the values ADD UP - just like resistors in SERIES. This way you can measure very small inductors. VERY CLEVER!

Question from a reader: *Can I add an inductor to stop a fuse blowing?*

Basically, an inductor NEVER prevents a fuse blowing because an inductor prevents spikes on one lead (we will call the INPUT lead), appearing on its other lead.
This is the detection and prevention of current that exists for a very short period of time.
A fuse detects an excess of current that occurs over a very long period of time and they are entirely two different "detectors."
One cannot assist the other in any way.

An inductor is basically a coil of wire. It may be thick or thin wire. The value of the inductor is a combination of the number of turns and the material on which the wire is wound.
The value of an inductor does not change over say a period of 20 years but it can go faulty by the enamel cracking and two turns touching. This can also be due to the difference in voltage between the two turns creating a spark between the turns and creating a "short."
When you test it, the high voltage is not present and it will test ok.
You may not think a few turns of wire will have any effect on improving a circuit, but spikes are very high frequency and the inductor will have a very big effect on reducing them.
An inductor (say 100uH) can be produced in many different sizes and the thickness of the wire will be important as it determines the current that can flow through the inductor.
The term "inductor" also includes those with two or more windings and these components are called TRANSFORMERS. These devices can get "shorts" and "leaks" between the windings and sparks can be seen between the windings. These sparks do not occur when you are testing them on test-equipment so the only way to

guarantee success is to replace it with an identical replacement.

TESTING SWITCHES and RELAYS
Switches and relays have contacts that open and close mechanically and you can test them for CONTINUITY. However these components can become intermittent due to dirt or pitting of the surface of the contacts due to arcing as the switch is opened.
It is best to test these items when the operating voltage and current is present as they quite often fail due to the arcing. A switch can work 49 times then fail on each 50th operation. The same with a relay. It can fail one time in 50 due to CONTACT WEAR.
If the contacts do not touch each other with a large amount of force and with a large amount of the metal touching, the current flowing through the contacts will create HEAT and this will damage the metal and sometimes reduce the pressure holding the contact together.
This causes more arcing and eventually the switch heats up and starts to burn.
Switches are the biggest causes of fire in electrical equipment and households.

A relay also has a set of contacts that can cause problems.
There are many different types of relays and basically they can be put into two groups.
1. An electromagnetic relay is a switch operated by magnetic force. This force is generated by current through a coil. The relay opens and closes a set of contacts. The contacts allow a current to flow and this current can damage the contacts.
Connect 5v or 12v to the coil (or 24v) and listen for the "click" of the points closing.
Measure the resistance across the points to see if they are closing.
You really need to put a load on the points to see if they are clean and can carry a current.
The coil will work in either direction.
If not, the relay is possibly a CMOS relay or Solid State relay.
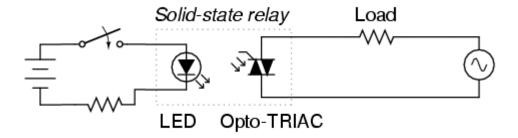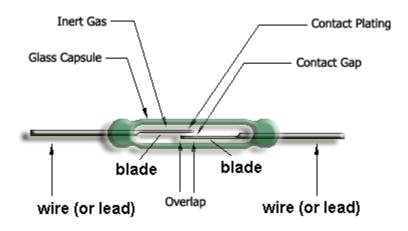
2. An electronic relay (Solid State Relay) does not have a winding. It works on the principle of an opto-coupler and uses a LED and Light Activated SCR or Opto-TRIAC to produce a low resistance on the output. The two pins that energise the relay (the two input pins) must be connected to 5v (or 12v) around the correct way as the voltage is driving a LED (with series resistor). The LED illuminates and activates a light-sensitive device.
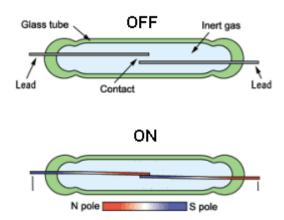
TESTING REED SWITCHES
A reed switch is generally contained in a long glass tube:

A wire or lead comes out each end for soldering to the reed switch to the project. The two "blades" inside the switch are made from a material that can be magnetised but does not retain its magnetism. This effect is called "temporally magnetised" (not permanently magnetised) and really only "passes" magnetism from one end to the other when in the presence of a magnet. One of the blades is made of a soft material and it will bend very easily. The other one is much stiffer.

When a magnet is placed under the two blades (or on top), the magnetism from the magnet is passed to the two blades (INDUCTION or MUTUAL INDUCTION - commonly called INDUCED) and it produces a very weak magnet (in the blade) that is identical to the powerful magnet as far as the position of the north and south poles are concerned). Initially it produces a N-S and N-S set of poles and this makes the two blades click together because the top blade will be South at the contact and the bottom blade will be North.
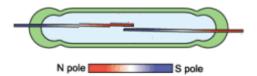
When the two blades click together the magnetism runs through the two blades and keeps them together. The two blades attract and the switch is closed. When the magnet is removed, the magnetism in the two blades ceases and the two blades move apart.

Since there is a very small amount of movement in the top blade, this switch has a limited number of operations. Eventually it will fail. It is a mechanical device and is not suited for detecting a spinning shaft as 100,000 revolutions will very quickly weaken the switch.
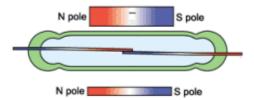
If the switch does not make contact or remains closed, the moveable blade can be cracked or broken. This can be very hard to see. So replace the switch.

LATCHING REED SWITCH
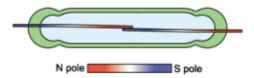
A "normal" reed switch can be converted into a **LATCHING REED SWITCH** by carefully placing a magnet below the switch and moving it away so the two blades open. Now move it slightly closer but do not allow the blades to close.
This is called putting a "SET" on the switch and the two blades will have a small magnetic effect "induced" in them but not enough to close the contacts:

N pole ▮▮▮▮▮ S pole
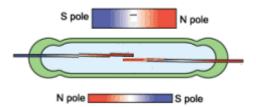
Now bring a strong magnet up to the reed switch on the other side of the glass tube with the north pole above the north of the lower magnet. This effect will increase the INDUCED MAGNETISM in the blades and close the contacts:
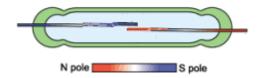
Remove the top magnet and the lower magnet will induce enough magnetism into the blades to keep them closed:

Now bring the upper magnet near the reed switch with the south pole above the north pole of the lower magnet. (In other words: AROUND THE OTHER WAY) This will have the effect of reducing the induced magnetism in the blades and a point will be reached when the two contacts will separate:

Remove the top magnet and the switch will remain separated because the lower magnet will not have sufficient influence on the blades to close the contact:

CAPACITORS

Capacitors are one of the most difficult things to test. That's because they don't give a reading on a multimeter and their value can range from 1p to 100,000u.
A faulty capacitor may be "open" when measured with a multimeter, and a good capacitor will also be "open."

You need a piece of test equipment called a CAPACITANCE METER to measure the value of a capacitor.

HOW A CAPACITOR WORKS
There are two ways to describe how a capacitor works. Both are correct and you have to combine them to get a full picture.
A capacitor has INFINITE resistance between one lead and the other.
This means no current flows **through** a capacitor. But it works in another way.
Suppose you have a strong magnet on one side of a door and a piece of metal on the other. By sliding the magnet up and down the door, the metal rises and falls.
The metal can be connected to a pump and you can pump water by sliding the magnet up and down.
A capacitor works in exactly the same way.
If you raise a voltage on one lead of a capacitor, the other lead will rise to the same voltage. This needs more explaining - we are keeping the discussion simple.
It works just like the magnetic field of the magnet through a door.
The next concept is this:
Capacitors are equivalent to a tiny rechargeable battery.
They store energy when the supply-voltage is present and release it when the supply drops.
These two concepts can be used in many ways and that's why capacitors perform tasks such as filtering, time-delays, passing a signal from one stage to another and create many different effects in a circuit.
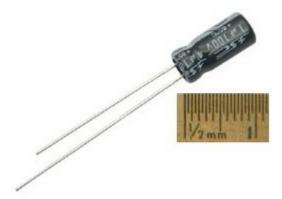
CAPACITOR VALUES
The basic unit of capacitance is the FARAD. (C) This is the value used in all equations, but it is a very large value. A one FARAD capacitor would be the size of a car if made with plates and paper. Most electronic circuits use capacitors with smaller values such as 1p to 1,000u. 1p is about equal to two parallel wires 2cm long. 1p is one picofarad.

The easiest way to understand capacitor values is to start with a value of 1u. This is one microfarad and is one-millionth of a Farad. A 1 microfarad capacitor is about 1cm long and the diagram shows a 1u electrolytic.

Smaller capacitors are ceramic and they look like the following. This is a 100n (0.1u)ceramic:

To read the value on a capacitor you need to know a few facts.

The basic value of capacitance is the FARAD.
1 microfarad is one millionth of 1 farad.
1 microfarad is divided into smaller parts called nanofarad.
1,000 nanofarad = 1 microfarad
Nanofarad is divided into small parts called picofarad
1,000 picofarad = 1 nanofarad.

Recapping:
1p = 1 picofarad.   1,000p = 1n (1 nanofarad)
1,000n = 1u (1 microfarad)
1,000u = 1millifarad
1,000,000u = 1 FARAD.

Examples:
All ceramic capacitors are marked in "p" (puff")
A ceramic with 22 is 22p = 22 picofarad
A ceramic with 47 is 47p = 47 picofarad
A ceramic with 470 is 470p = 470 picofarad
A ceramic with 471 is 470p = 470 picofarad
A ceramic with 102 is 1,000p = 1n
A ceramic with 223 is 22,000p = 22n
A ceramic with 104 is 100,000p = 100n = 0.1u

TYPES OF CAPACITOR
For testing purposes, there are two types of capacitor.
Capacitors from 1p to 100n are non-polar and can be inserted into a circuit around either way.
Capacitors from 1u to 100,000u are electrolytics and are polarised. They must be fitted so the positive lead goes to the supply voltage and the negative lead goes to ground (or earth).
There are many different sizes, shapes and types of capacitor. They are all the same. They consist of two plates with an insulating material between. The two plates can be stacked in layers or rolled together.
The important factor is the insulating material. It must be very thin to keep things small. This gives the capacitor its VOLTAGE RATING.
If a capacitor sees a voltage higher than its rating, the voltage will "jump through" the insulating material or around it.
If this happens, a carbon deposit is left behind and the capacitor becomes "leaky" or very low resistance, as carbon is conductive.
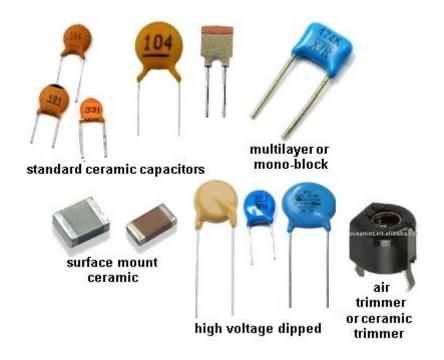
CERAMIC CAPACITORS

Nearly all small capacitors are **ceramic capacitors** as this material is cheap and the capacitor can be made in very thin layers to produced a high capacitance for the size of the component. This is especially true for surface-mount capacitors.
All capacitors are marked with a value and the basic unit is: "p" for "puff" However NO surface mount capacitors are marked and they are very difficult to test.

standard ceramic capacitors

multilayer or
mono-block

surface mount
ceramic

high voltage dipped

air
trimmer
or ceramic
trimmer
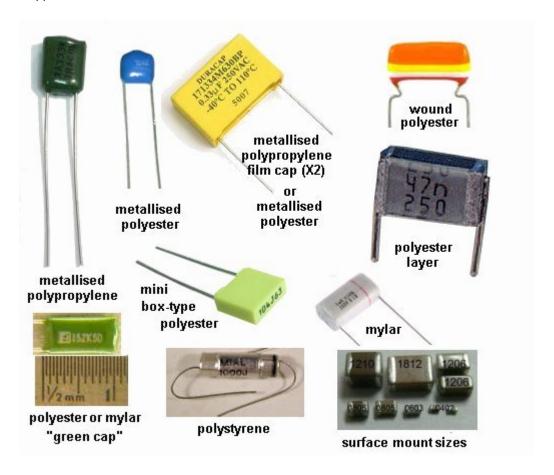
VALUE:	VALUE WRITTEN ON THE COMPONENT:
0.1p	0p1
0.22p	0p22
0.47p	0p47
1.0p	1p0
2.2p	2p2
4.7p	4p7
5.6p	5p6
8.2p	8p2
10p	10 or 10p
22p	22 or 22p
47p	47 or 47p
56p	56 or 56p
100p	100 on 101
220p	220 or 221
470p	470 or 471
560p	560 or 561
820p	820 or 821
1,000p (1n)	102
2200p (2n2)	222
4700p (4n7)	472
8200p (8n2)	822
10n	103
22n	223
47n	473
100n	104
220n	224
470n	474
1u	105

POLYESTER, POLYCARBONATE, POLYSTYRENE, MYLAR, METALLISED POLYESTER, ("POLY"), MICA and other types of CAPACITOR

There are many types of capacitor and they are chosen for their reliability, stability, temperate-range and cost.

For testing and repair work, they are all the same. Simply replace with exactly the same type and value.
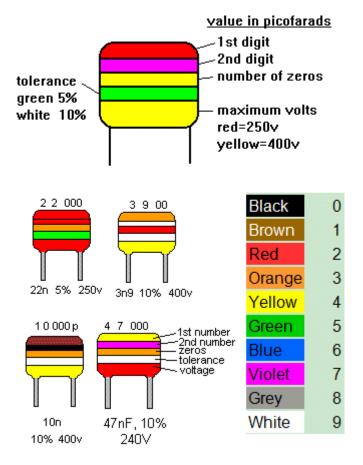
Capacitor Colour Code Table

Colour	Digit A	Digit B	Multiplier D	Tolerance (T) > 10pf	Tolerance (T) < 10pf	Temperature Coefficient (TC)
Black	0	0	x1	± 20%	± 2.0pF	
Brown	1	1	x10	± 1%	± 0.1pF	-33×10^{-6}
Red	2	2	x100	± 2%	± 0.25pF	-75×10^{-6}
Orange	3	3	x1,000	± 3%		-150×10^{-6}
Yellow	4	4	x10,000	± 4%		-220×10^{-6}
Green	5	5	x100,000	± 5%	± 0.5pF	-330×10^{-6}
Blue	6	6	x1,000,000			-470×10^{-6}
Violet	7	7				-750×10^{-6}
Grey	8	8	x0.01	+80%,-20%		
White	9	9	x0.1	± 10%	± 1.0pF	
Gold			x0.1	± 5%		

Silver			x0.01	± 10%		

Pico Farads (pF)	Nano Farads (nF)	Micro Farads (µF)
1	0.001	0.000001
10	0.01	0.00001
100	0.1	0.0001
1,000	1	0.001
10,000	10	0.01
100,000	100	0.1
1,000,000	1,000	1
10,000,000	10,000	10
100,000,000	100,000	100

Type ⊕ = polarized	Pic	Cap Range
Ceramic		pF - µF
Mica (silver mica)		pF - nF
Plastic Film (polyethylene polystyrene)		few µFs
Tantalum ⊕		µFs
OSCON ⊕		µFs
Aluminum Electrolytic ⊕		high µFs

value in picofarads
- 1st digit
- 2nd digit
- number of zeros
- maximum volts red=250v yellow=400v

tolerance green 5% white 10%

2 2 000 — 22n 5% 250v
3 9 00 — 3n9 10% 400v
1 0 000 p — 10n 10% 400v
4 7 000 — 47nF, 10% 240V
- 1st number
- 2nd number
- zeros
- tolerance
- voltage

Color	Value
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Grey	8
White	9

ELECTROLYTIC and TANTALUM CAPACITORS
Electrolytics and Tantalums are the same for testing purposes but their performance is slightly different in some circuits. A tantalum is smaller for the same rating as an electrolytic and has a better ability at delivering a current. They are available up to about 1,000u, at about 50v but their cost is much higher than an electrolytic.
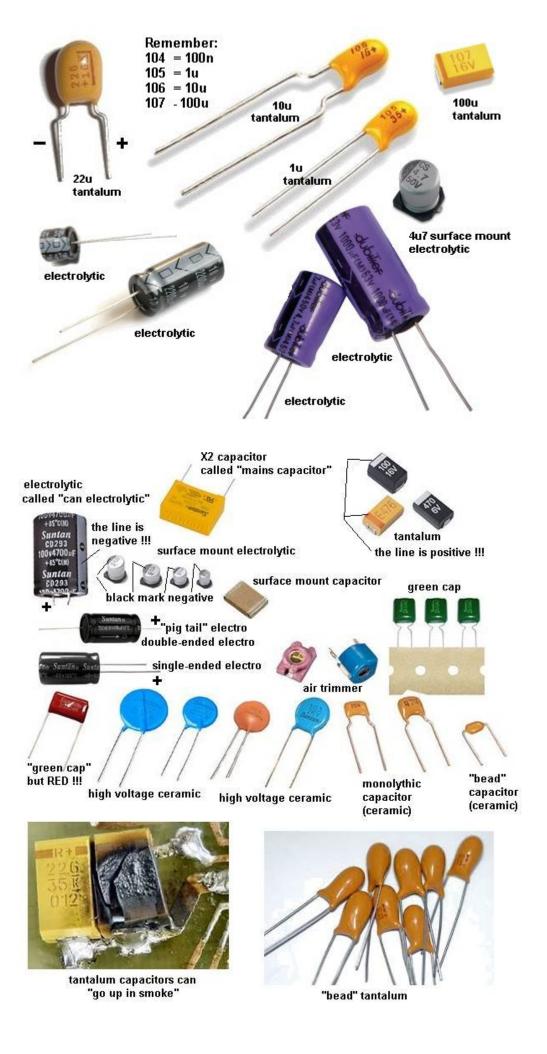
Electrolytics are available in 1u, 2u2 3u3 4u7 10u, 22u, 47u, 100u, 220u, 330u, 470u, 1,000u, 2,200u, 3,300u, 4,700u, 10,000u and higher.
The "voltage" or "working voltage" can be: 3.3v, 10v, 16v, 25v, 63v, 100v, 200v and higher.
There is also another important factor that is rarely covered in text books. It is RIPPLE FACTOR.
This is the amount of current that can enter and leave an electrolytic. This current heats up the electrolytic and that is why some electrolytics are much larger than others, even though the capacitance and voltage-ratings are the same.
If you replace an electrolytic with a "miniature" version, it will heat up and have a very short life. This is especially important in power supplies where current (energy) is constantly entering and exiting the electrolytic as its main purpose is to provide a smooth output from a set of diodes that delivers "pulsing DC." (see "Power Diodes")

Remember:
104 = 100n
105 = 1u
106 = 10u
107 - 100u

22u tantalum

10u tantalum

1u tantalum

100u tantalum

4u7 surface mount electrolytic

electrolytic

electrolytic

electrolytic

electrolytic

electrolytic
called "can electrolytic"

the line is negative !!!

X2 capacitor
called "mains capacitor"

surface mount electrolytic

tantalum
the line is positive !!!

black mark negative

surface mount capacitor

green cap

"pig tail" electro
double-ended electro

single-ended electro

air trimmer

"green cap"
but RED !!!

high voltage ceramic

high voltage ceramic

monolythic capacitor (ceramic)

"bead" capacitor (ceramic)

tantalum capacitors can
"go up in smoke"

"bead" tantalum

PARALLEL and SERIES CAPACITORS

Capacitors can be connected in PARALLEL and/or SERIES for a number of reasons.
1. If you do not have the exact value, two or more connected in parallel or series can produce the value you need.
2. Capacitors connected in series will produce one with a higher voltage rating.
3. Capacitors connected in parallel will produce a larger-value capacitance.

Here are examples of two equal capacitors connected in series or parallel and the results they produce:

NON-POLAR CAPACITORS (ELECTROLYTICS)
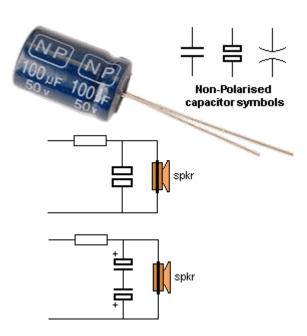
Electrolytics are also available in non-polar values. It sometimes has the letters "NP" on the component. Sometimes the leads are not identified.
This is an electrolytic that does not have a positive and negative lead but two leads and either lead can be connected to the positive or negative of the circuit.
These electrolytics are usually connected to the output of an amplifier (such as in a filter near the speaker) where the signal is rising and falling.
A non-polar electrolytic can be created from two ordinary electrolytics by connecting the negative leads together and the two positive leads become the new leads.
For example: two 100u 63v electrolytics will produce a 47u 63v non-polar electrolytic.
In the circuit below, the non-polar capacitor is replaced with two electrolytics.
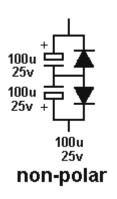
MAKING A NON-POLAR ELECTROLYTIC

non-polar

A normal electrolytic must be connected the correct way in a circuit because it has a thin insulating layer covering the plates that has a high resistance.
If you connect the electrolytic around the wrong way, this layer "breaks-down" and the resistance of the electrolytic becomes very small and a high current flows. This heats up the electrolytic and the current increases. Very soon the capacitor produces gasses and explodes.
One big mistake in many text books shows how to make a non-polar electrolytic by connecting two "back-to-back."
They claim 2 x 100u connected back-to-back is equal to 47u.
This appears to be case when testing on a meter but the meter simply charges them for a short period of time to get a reading.
If you allow them to charge fully you will find the reverse electrolytic has a very small voltage across it.
Secondly, when you are charging them, you are putting a high current through the reverse electrolytic and damaging the layer.
To prevent this, you need to add two diodes as shown in the diagram.
In addition, 2 x 100u "back-to-back" is very near 100u.

Here is a question from a reader:
I have an amplifier with 2 x2,200u electrolytics on the output of a bridge. Can I replace them with a single 10,000u?

You need to look at the circuit of your amplifier. The two 2,200u electrolytics are possibly connected as shown in the circuit above and you will notice they are joined to produce a positive rail and a negative rail with zero (called earth) in the centre.
This forms two different circuits with the top electrolytic filtering the positive rail and the bottom electro filtering the negative rail. They must be connected to the zero volts rail.
A single 10,000u cannot be connected to the 0v rail and cannot be substituted for the two electro's.
You can easily determine of the two electro's are connected as shown above.
Test the positive terminal of each electro by placing the negative of the meter on the chassis.
If the positive of one electro have zero volts, it will be the lower electro in the diagram above. The negative terminal of the electro will have a minus voltage on it.

VOLTAGE RATING OF CAPACITOR
Capacitors have a voltage rating, stated as WV for working voltage, or WVDC. This specifies the maximum voltage that can be applied across the capacitor without puncturing the dielectric. Voltage ratings for "poly," mica and ceramic capacitors are typically 50v to 500 VDC. Ceramic capacitors with ratings of 1kv to 5kv are also available. Electrolytic capacitors are commonly available in 6v, 10v 16v, 25v, 50v, 100v, 150v, and 450v ratings.

THE SIZE OF A CAPACITOR - RIPPLE FACTOR

The size of a capacitor depends on a number of factors, namely the value of the capacitor (in microfarads etc) and the voltage rating. But there is also another factor that is most important. It is the RIPPLE FACTOR. **Ripple Factor** is the amount of voltage-fluctuation the capacitor (electrolytic) can withstand without getting too hot. When current flows in and out of an electrolytic, it gets hot and this will eventually dry-out the capacitor as some of the liquid inside the capacitor escapes through the seal. It's a very slow process but over a period of years, the capacitor looses its capacitance.

If you have two identical 1,000u 35v electrolytics and one is smaller, it will get hotter when operating in a circuit and that's why it is necessary to choose the largest electrolytic.

CAUTION

If a capacitor has a voltage rating of 63v, do not put it in a 100v circuit as the insulation (called the dielectric) will be punctured and the capacitor will "short-circuit." It's ok to replace a 0.22uF 50WV capacitor with 0.22uF 250WVDC.

SAFETY

A capacitor can store a charge for a period of time after the equipment is turned off. High voltage electrolytic caps can pose a safety hazard. These capacitors are in power supplies and some have a resistor across them, called a bleed resistor, to discharge the cap after power is switched off.

If a bleed resistor is not present the cap can retain a charge after the equipment is unplugged.

How to discharge a capacitor

Do not use a screwdriver to short between the terminals as this will damage the capacitor internally and the screwdriver.

Use a 1k 1 watt or 3watt or 5watt resistor on jumper leads (or held with pliers) and keep them connected for up to 15 seconds to fully discharge the electro.

Test it with a voltmeter to make sure all the energy has been removed.

Before testing any capacitors, especially electrolytics, you should look to see if any are damaged, overheated or leaking. Swelling at the top of an electrolytic indicates heating (and pressure inside the case) and will result in drying out of the electrolyte. Any hot or warm electrolytic indicates leakage and ceramic capacitors with portions missing indicates something has gone wrong (such as it being "blown apart").

Here is a 120u 330v electrolytic from a flash circuit in an old-fashioned film camera. If the flash does not "fire," the electrolytic will be charged to about 350 volts !! Use a 1k resistor (held with pliers) to slowly discharge it. It may take 15 seconds to fully discharge.

TESTING A CAPACITOR
There are two things you can test with a multimeter:
1. A short-circuit within the capacitor
2. Capacitor values above 1u.

You can test capacitors in-circuit for short-circuits. Use the x1 ohms range.
To test a capacitor for leakage, you need to remove it or at least one lead must be removed. Use the x10k range on an analogue or digital multimeter.
For values above 1u you can determine if the capacitor is charging by using an analogue meter. The needle will initially move across the scale to indicate the cap is charging, then go to "no deflection." Any permanent deflection of the needle will indicate leakage.
You can reverse the probes to see if the needle moves in the opposite direction. This indicates it has been charged. Values below 1u will not respond to charging and the needle will not deflect.
This does not work with a digital meter as the resistance range does not output any current and the electrolytic does not charge.

Rather than spending money on a capacitance meter, it is cheaper to replace any suspect capacitor or electrolytic.
Capacitors can produce very unusual faults and no piece of test equipment is going to detect the problem.
In most cases, it is a simple matter to solder another capacitor across the suspect component and view or listen to the result.
This saves all the worry of removing the component and testing it with equipment that cannot possibly give you an accurate reading when the full voltage and current is not present.
It is complete madness to even think of testing critical components such as capacitors, with TEST EQUIPMENT. You are fooling yourself. If the Test Equipment says the component is ok, you will look somewhere else and waste a lot of time.
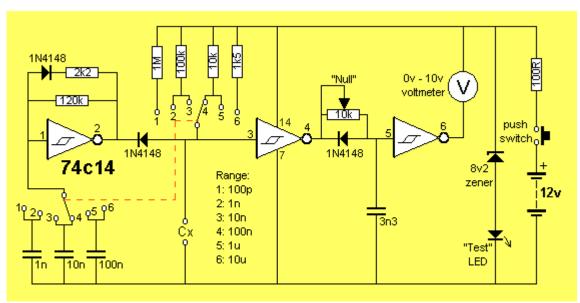
FINDING THE VALUE OF A CAPACITOR
If you want to find the value of a surface-mount capacitor or one where the markings have been removed, you will need a CAPACITANCE METER. Here is a simple circuit that can be added to your meter to read capacitor values from 10p to 10u.
The full article can be found HERE.

ADD-ON CAPACITANCE METER

CAPACITOR SUBSTITUTION BOX

You can get a kit or a ready-made piece of test gear called **CAPACITOR SUBSTITUTION BOX** and also **RESISTOR SUBSTITUTION BOX**.
I bought one of each 30 years ago and I have only used them ONCE.
They appear to be very handy but when you are testing a circuit, you want the component next to the other parts.
It is just as easy to pick the component you need from your junk box and connect it to the circuit via jumper leads.

REPLACING A CAPACITOR

Always replace a capacitor with the exact same type.
A capacitor may be slightly important in a circuit or it might be extremely critical.
A manufacturer may have taken years to select the right type of capacitor due to previous failures.
A capacitor just doesn't have a "value of capacitance."
It may also has an effect called "tightening of the rails."
In other words, a capacitor has the ability to react quickly and either absorb or deliver energy to prevent spikes or fluctuations on the rail.
This is due to the way it is constructed. Some capacitors are simply plates of metal film while others are wound in a coil. Some capacitors are large while others are small.
They all react differently when the voltage fluctuates.
Not only this, but some capacitors are very stable and all these features go into the decision for the type of capacitor to use.
You can completely destroy the operation of a circuit by selecting the wrong type of capacitor.
No capacitor is perfect and when it gets charged or discharged, it appears to have a small value of resistance in series with the value of capacitance. This is known as "ESR" and stands for EQUIVALENT SERIES RESISTANCE. This effectively makes the capacitor slightly slower to charge and discharge.
We cannot go into the theory on selecting a capacitor as it would be larger than this eBook so the only solution is to replace a capacitor with an identical type.
However if you get more than one repair with identical faults, you should ask other technicians if the original capacitor comes from a faulty batch.
The author has fixed TV's and fax machines where the capacitors have been inferior and alternate types have solved the problem.
Some capacitor are suitable for high frequencies, others for low frequencies.

DECOUPLING CAPACITORS

A Decoupling Capacitor can severe one, two or three functions. You need to think of a decoupling capacitor as a miniature battery with the ability to deliver a brief pulse of energy when ever the line-voltage drops and also absorb a brief pulse of energy when ever the line voltage rises (or spikes).
Decoupling capacitor can range from 100n to 1,000u.
100n capacitors are designed to absorb spikes and also have the effect of tightening-

up the rails for high frequencies. They have no effect on low frequencies such as audio frequencies.

These capacitors are generally ceramic and have very low internal impedance and thus they can operate at high frequencies.

Capacitors above about 10u are used for decoupling and these are nearly always electrolytics.

Decoupling means "tightening-up the power rails." The electrolytic acts just like a miniature rechargeable battery, supplying a small number of components in a circuit with a smooth and stable voltage.

The electrolytic is usually fed from a dropper resistor and this resistor charges the electrolytic and adds to the ability of the electrolytic to create a "separate power supply."

These two components help remove spikes as an electrolytic cannot remove spikes if connected directly to the supply rails - it's internal impedance is high and the spikes are not absorbed.

Decoupling capacitors are very difficult to test.

They rarely fail but if a project is suffering from unknown glitches and spikes, it is best to simply add more 100n decoupling caps on the underside of the board and replace all electrolytics.

Some small electrolytics will dry out due to faulty manufacture and simply replacing every one on a board will solve the problem.

Some of the functions of a decoupling capacitor are:

Removing ripple - hum or buzz in the background of an amplifier

Removing glitches or spikes.

Separating one stage from another to reduce or remove MOTORBOATING - a low frequency sound due to the output putting a pulse on the power rails that is picked up by the pre-amplifier section and amplified.

TESTING DIODES

Diodes can have 4 different faults.
1. Open circuit in both directions.
2. Low resistance in both directions.
3. Leaky.
4. Breakdown under load.

TESTING A DIODE ON AN ANALOGUE METER

Testing a diode with an **Analogue Multimeter** can be done on any of the resistance ranges. [The high resistance range is best - it sometimes has a high voltage battery for this range but this does not affect our testing]

There are two things you must remember.

1. When the diode is measured in one direction, the needle will **not move at all**. The technical term for this is the diode is **reverse biased**. It will not allow any current to flow. Thus the needle will not move.

When the diode is connected around the other way, the needle will swing to the right (move up scale) to about 80% of the scale. This position represents the voltage drop across the junction of the diode and is NOT a resistance value. If you change the resistance range, the needle will move to a slightly different position due to the resistances inside the meter. The technical term for this is the diode is **forward biased**. This indicates the diode is not faulty.

The needle will swing to a slightly different position for a "normal diode" compared to a Schottky diode. This is due to the different junction voltage drops.

However we are only testing the diode at very low voltage and it may break-down when fitted to a circuit due to a higher voltage being present or due to a high current flowing.

2. The leads of an **Analogue Multimeter** have the positive of the battery connected to the black probe and the readings of a "good diode" are shown in the following two diagrams:

**The diode is REVERSE BIASED in the
diagram above and diodes not conduct.**

Here is the equivalent circuit:

**The diode is FORWARD BIASED in the
diagram above and it conducts**

TESTING A DIODE ON A DIGITAL METER
Testing a diode with a Digital Meter must be done on the "DIODE" setting as a digital
meter does not deliver a current through the probes on some of the resistance
settings and will not produce an accurate reading.

The best thing to do with a "suspect" diode is to replace it. This is because a diode
has a number of characteristics that cannot be tested with simple equipment. Some
diodes have a fast recovery for use in high frequency circuits. They conduct very
quickly and turn off very quickly so the waveform is processed accurately and

efficiently.

If the diode is replaced with an ordinary diode, it will heat up as does not have the high-speed characteristic.

Other diodes have a low drop across them and if an ordinary is used, it will heat up.

Most diodes fail by going: SHORT-CIRCUIT. This can be detected by a low resistance (x1 or x10 Ohms range) in both directions.

A diode can also go OPEN CIRCUIT. To locate this fault, place an identical diode across the diode being tested.

A leaky diode can be detected by a low reading in one direction and a slight reading the other direction.
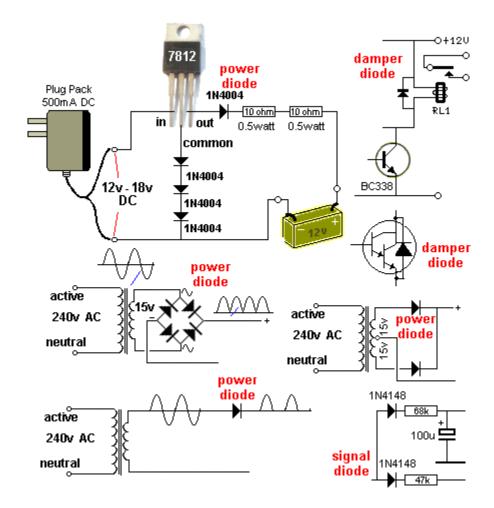
However this type of fault can only be detected when the circuit is working. The output of the circuit will be low and sometimes the diode heats up (more than normal).

A diode can go open under full load conditions and perform intermittently.

Diodes come in pairs in surface-mount packages and 4 diodes can be found in a bridge.

They are also available in pairs that look like a 3-leaded transistor.

The line on the end of the body of a diode indicates the cathode and you cannot say "this is the positive lead." The correct way to describe the leads is to say the "cathode lead." The other lead is the anode. The cathode is defined as the electrode (or lead) through which an electric current flows out of a device.

The following diagrams show different types of diodes:

POWER DIODES

To understand how a power diode works, we need to describe a few things. This has NEVER been described before, so read carefully.

The 240v AC (called the "mains") consists of two wires, one is called the ACTIVE and the other is NEUTRAL. Suppose you touch both wires. You will get a shock. The neutral is connected to an earth wire (or rod driven into the ground or connected to a water pipe) at the point where the electricity enters the premises and you do not get a shock from the NEUTRAL.

But the voltage on the active is rising to +345v then goes to -345v at the rate of 50 times per second (for a complete cycle).
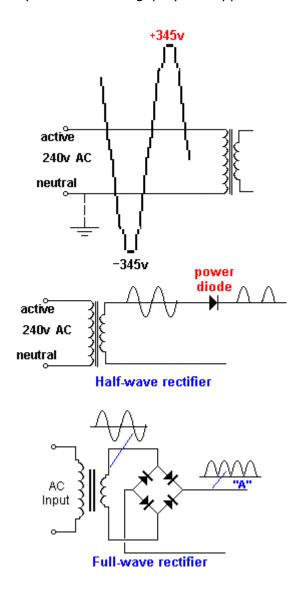
345v is the peak voltage of 240v. You never get a 240v shock. (It is a 345v shock.) In other words, if you touch the two wires at a particular instant, you would get a POSITIVE 345v shock and at another instant you would get a negative 345v shock. This is shown in the diagram below.

We now transfer this concept to the output of a transformer. The diagram shows an AC waveform on the output of the secondary.

This voltage is rising 15v higher than the bottom lead then it is 15v LOWER than the bottom lead. The bottom lead is called "zero volts." You have to say one lead or wire is not "rising and falling" as you need a "reference" or starting-point" or "zero point" for voltage measurements.

The diode only conducts when the voltage is "above zero" (actually when it is 0.7v above zero) and does not conduct (at all) when the voltage goes below zero.

This is shown on the output of the Power Diode. Only the positive peaks or the positive parts of the waveform appear on the output and this is called "pulsing DC." This is called "half-wave" and is not used in a power supply. We have used it to describe how the diode works. The electrolytics charge during the peaks and deliver energy when the diode is not delivering current. This is how the output becomes a steady DC voltage.

Power supplies use FULL WAVE rectification and the other half of the AC waveform is delivered to the output (and fills in the "gaps") and appears as shown in "A."

Half-wave rectifier

Full-wave rectifier
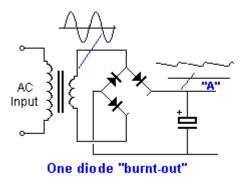
ONE FAULTY DIODE

One diode in a bridge can go open (any of the 4 diodes will produce the same effect) and produce an output voltage that can be slightly lower than the original voltage.

The actual "voltage-drop" will depend on the current taken by the circuit and the ability of the transformer to produce the required voltage and current during half-wave operation. The voltage during each half cycle (when none of the diodes is delivering any energy to the circuit) is maintained by the electrolytic and its size (relative to the current taken by the circuit) will determine the size of the ripple that will result when the diode fails. The ripple will be 100 to 1,000 times greater after the failure of a diode, depending on the value of the filter capacitor.

To locate the faulty diode, simply get a diode and place it across each of the diodes in the bridge (in turn) when the circuit is working.

For a bridge rectifier, the ripple-frequency will be twice the mains frequency and its ripple will be very small if the electrolytic is the correct value. When a diode fails, the ripple-frequency will be equal to mains-frequency and the amplitude will increase considerably. You may even hear background hum from audio equipment.

If you cannot find a faulty diode, the filter capacitor will be at fault. Turn off the equipment and connect an electrolytic across the filter capacitor via jumper leads. Turn the power ON and see if the hum has reduced.

One diode "burnt-out"

DAMPER DIODES

A damper diode is a diode that detects a high voltage and SQUELCHES IT (reduces it - removes it). The signal that it squelches is a voltage that is in the opposite direction to the "supply voltage" and is produced by the collapsing of a magnetic field. Whenever a magnetic filed collapses, it produces a voltage in the winding that is opposite to the supply voltage and can be much higher. This is the principle of a flyback circuit or EHT circuit. The high voltage comes from the transformer.

The diode is placed so that the signal passes through it and less than 0.5v appears across it.

A damper diode can be placed across the coil of a relay, incorporated into a transistor or FET or placed across a winding of a flyback transformer to protect the driving transistor or FET.

It can also be called a "Reverse-Voltage Protection Diode," "Spike Suppression Diode," or "Voltage Clamp Diode."

The main characteristic of a Damper Diode is HIGH SPEED so it can detect the spike and absorb the energy.

It does not have to be a high-voltage diode as the high voltage in the circuit is being absorbed by the diode.

SILICON, GERMANIUM AND SCHOTTKY DIODES

When testing a diode with an analogue meter, you will get a low reading in one direction and a high (or NO READING) in the other direction. When reading in the LOW direction, the needle will swing nearly full scale and the reading is not a resistance-value but a reflection of the characteristic voltage drop across the junction of the diode. As we mentioned before, a resistance reading is really a voltage reading and the meter is measuring the voltage of the battery minus the voltage-drop across the diode.

Since Silicon, Germanium and Schottky Diodes have slightly different characteristic voltage drops across the junction, you will get a slightly different reading on the scale. This does not represent one diode being better than the other or capable of handling a higher current or any other feature.

The quickest, easiest and cheapest way to find, fix and solve a problem caused by a

faulty diode is to replace it.
There is no piece of test equipment capable of testing a diode fully, and the circuit you are working on is actually the best piece of test equipment as it is identifying the fault UNDER LOAD.

Only very simple tests can be done with a multimeter and it is best to check a diode with an ANALOGUE MULTIMETER as it outputs a higher current though the diode and produces a more-reliable result.
A Digital meter can produce false readings as it does not apply enough current to activate the junction.
Fortunately almost every digital multimeter has a **diode test mode.** Using this, a silicon diode should read a voltage drop between 0.5v to 0.8v in the forward direction and open in the reverse direction. For a germanium diode, the reading will be lower, around 0.2v - 0.4v in the forward direction. A bad diode will read zero volts in both directions.

REPLACING A DIODE
It is alway best to replace a diode with the same type but quite often this is not possible. Many diodes have unusual markings or colours or "in-house" letters.
This is only a general guide because many diodes have special features, especially when used in high-frequency circuits.
However if you are desperate to get a piece of equipment working, here are the steps:
Determine if the diode is a signal diode, power diode, or zener diode.
For a signal diode, try 1N4148.
For a power diode (1 amp) try 1N4004. (for up to 400v)
For a power diode (3 amp) try 1N5404. (for up to 400v)
For a high-speed diode, try UF4004 (for up to 400v)

If you put an ordinary diode in a high-speed application, it will get very hot very quickly.
To replace an unknown zener diode, start with a low voltage such as 6v2 and see if the circuit works.
The size of a diode and the thickness of the leads will give an idea of the current-capability of the diode.
Keep the leads short as the PC board acts as a heat-sink.
You can also add fins to the leads to keep the diode cool.

LIGHT EMITTING DIODES (LEDs)
Light Emitting Diodes (LEDs) are diodes that produce light when current flows from anode to cathode. The LED does not emit light when it is revered-biased. It is used as a low current indicator in many types of consumer and industrial equipment, such as monitors, TV's, printers, hi-fi systems, machinery and control panels.
The light produced by a LED can be visible, such as red, green, yellow or white. It can also be invisible and these LEDs are called Infrared LEDs. They are used in remote controls and to see if they are working, you need to point a digital camera at the LED and view the picture on the camera screen.
An LED needs about 2v - 3.6v across its leads to make it emit light, but this voltage must be exact for the type and colour of the LED. The simplest way to deliver the exact voltage is to have a supply that is higher than needed and include a voltage-dropping resistor. The value of the resistor must be selected so the current is between 2mA and 25mA.
The cathode of the LED is identified by a flat on the side of the LED. The life expectancy of a LED is about 100,000 hours. LEDs rarely fail but they are very sensitive to heat and they must be soldered and de-soldered quickly. They are one of the most heat-sensitive components.
Light emitting diodes cannot be tested with most multimeters because the characteristic voltage across them is higher than the voltage of the battery in the meter.
However a simple tester can be made by joining 3 cells together with a 220R resistor
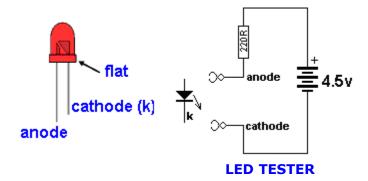
and 2 alligator clips:

LED TESTER

Connect the clips to a LED and it will illuminate in only one direction.
The colour of the LED will determine the voltage across it. You can measure this voltage if you want to match two or more LEDs for identical operation.
Red LEDs are generally 1.7v to 1.9v. - depending on the quality such as "high-bright"
Green LEDs are 1.9v to 2.3v.
Orange LEDs are about 2.3v and
White LEDs and IR LEDs are about 3.3v to 3.6v.
The illumination produced by a LED is determined by the quality of the crystal. It is the crystal that produces the colour and you need to replace a LED with the same quality to achieve the same illumination.
Never connect a LED across a battery (such as 6v or 9v), as it will be instantly damaged. You must have a resistor in series with the LED to limit the current.

ZENER DIODES

All diodes are Zener diodes. For instance a 1N4148 is a 120v zener diode as this is its reverse breakdown voltage.
And a zener diode can be used as an ordinary diode in a circuit with a voltage that is below the zener value.
For instance, 20v zener diodes can be used in a 12v power supply as the voltage never reaches 20v, and the zener characteristic is never reached.
Most diodes have a reverse breakdown voltage above 100v, while most zeners are below 70v. A 24v zener can be created by using two 12v zeners in series and a normal diode has a characteristic voltage of 0.7v. This can be used to increase the voltage of a zener diode by 0.7v. See the diagram above. It uses 3 ordinary diodes to increase the output voltage of a 3-terminal regulator by 2.1v.
To tests a zener diode you need a power supply about 10v higher than the zener of the diode. Connect the zener across the supply with a 1k to 4k7 resistor and measure the voltage across the diode. If it measures less than 1v, reverse the zener. If the reading is high or low in both directions, the zener is damaged.
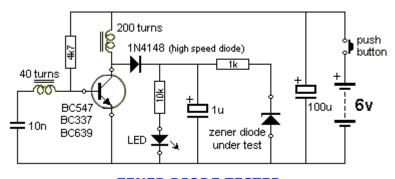
Here is a zener diode tester. The circuit will test up to 56v zeners.
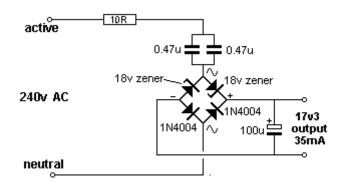
ZENER DIODE TESTER

TRANSFORMERLESS POWER SUPPLY
Here's a circuit that uses zener diodes in a power supply to show how they work.
This clever design uses 4 diodes in a bridge to produce a fixed voltage power supply
capable of supplying 35mA.
If we put 2 zener diodes in a bridge with two ordinary power diodes, the bridge will
break-down at the voltage of the zener. This is what we have done. If we use 18v
zeners, the output will be 17v4.

SUPPLY USING ZENER DIODES

When the incoming voltage is positive at the top, the left zener provides 18v limit
(and the other zener produces a drop of 0.6v). This allows the right zener to pass
current just like a normal diode. The output is 17v4. The same with the other half-
cycle.
You cannot use this type of bridge in a normal power supply as the zener diode will
"short" when the input voltage reaches the zener value. The concept only works in
the circuit above.

VOLTAGE REGULATORS
A Voltage Regulator takes a high input voltage and delivers a fixed output voltage.
Providing the input voltage is 4v above the output voltage, the regulator will deliver
a fixed output voltage with almost no ripple.
Voltage regulators are also called "3-TERMINAL REGULATORS" or "REGULATOR IC's"
- although this name is not generally used.
In most cases, a voltage regulator gets quite hot and for this reason it has a high
failure-rate.
If a regulator is not getting hot (or warm) it has either failed or the circuit is not
operating.
A regulator can only decrease the voltage. It cannot increase the current. This
means the current being supplied to a circuit must also be available from the circuit
supplying the regulator.
All regulators have different pin-outs, so you need to find the input pin and output
pin and make sure the voltage-difference is at least 4v. Some regulators will work
with a difference as low as 1v, so you need to read the specifications for the type
you are servicing.
Some regulators are called "negative voltage regulators" and the input voltage will
be negative and the output will be negative.
You need to test a voltage regulator with the power "ON".
Make sure you do not allow the probes to short any of the pins together as this will
destroy the regulator or the circuit being supplied.
With the power turned off or the regulator removed from the circuit, you can test it
with a multimeter set to resistance to see if it is ok. If any resistance readings are
very low or zero ohms, the regulator is damaged.

TRANSFORMERS
All transformers and coils are tested the same way. This includes chokes, coils,

inductors, yokes, power transformers, EHT transformers (flyback transformers), switch mode transformers, isolation transformers, IF transformers, baluns, and any device that has turns of wire around a former. All these devices can go faulty.
The coating on the wire is called insulation or "enamel" and this can crack or become overheated or damaged due to vibration or movement. When two turns touch each other, a very interesting thing happens. **The winding becomes two separate windings**.

We will take the case of a single winding such as a coil. This is shown in the first diagram above and the winding is wound across a former (a former is a bobbin or plastic molding or something to hold the winding) and back again, making two layers. The bottom and top layers touch at the point shown in the diagram and the current that originally passed though A, B, C, D now passes though A & D.

Winding B C becomes a separate winding as shown in the second diagram.
In other words the coil becomes a TRANSFORMER with a SHORT CIRCUIT on the secondary winding as shown in the third diagram.
When the output wires of a transformer are shorted together, it delivers a very high current because you have created a SHORT-CIRCUIT. This short-circuit causes the transformer to get very hot.
That's exactly what happens when any coil or transformer gets a "shorted turn."
The shorted turns can be a single turn or many turns.
It is not possible to measure a fault like this with a multimeter as you don't know the exact resistance of a working coil or winding and the resistance of a faulty winding may be only 0.001 ohms less.
However when a transformer or coil is measured with an inductance meter, an oscillating voltage (or spike) is delivered into the core as magnetic flux, then the magnetic flux collapses and passes the energy into the winding to produce a waveform. The inductance meter reads this and produces a value of inductance in Henry (or milliHenry or microHenry.)
This is done with the transformer removed from the circuit and this can be a very difficult thing to do, as most transformers have a number of connections.
If the coil or transformer has a shorted turn, the energy from the magnetic flux will pass into the turns that are shorted and produce a current. Almost no voltage will be detected from winding.
The reading from the inductance meter will be low or very low and you have to work out if it is correct.
However there is one major problem with measuring a faulty transformer or coil.
It may only become faulty when power is applied.
The voltage between the turns may be sparking or jumping a gap and creating a problem. A tester is not going to find this fault.
Secondly, an inductance meter may produce a reading but you do not know if the reading is correct. An improved tester is a RING TESTER.
The circuit for a ring tester can be found here:

http://www.flippers.com/pdfs/k7205.pdf

It sends a pulse to the coil and counts the number of returning pulses or "rings." A faulty coil (or winding) may return one pulse but nearly all the energy will be passed to the shorted turns and you will be able to see this on the scale. You will only get one or two return pulses, whereas a good winding will return more pulses.

One way to detect a faulty power transformer is to connect it to the supply and feel the temperature-rise (when nothing is connected to the secondary).
It should NOT get hot.
Detecting shorted turns is not easy to diagnose as you really need another identical

component to compare the results.

Most transformers get very hot when a shorted turn has developed. It may deliver a voltage but the heat generated and a smell from the transformer will indicate a fault.

ISOLATION TRANSFORMER

An isolation transformer is a piece of **Test Equipment** that provides "Mains Voltage" but the voltage is "floating." You will still get a shock if you touch the two output leads, but it has a special use when testing unknown equipment.

Many electrical appliances are fully insulated and only have two leads connected to the mains.

When you take these appliances apart, you do not know which end of say a heating element is connected to the "live" (active) side of the mains and which end connects to the neutral.

I am not suggesting you carry out the following tests, but they are described to show how an isolation transformer works.

If you touch a soldering iron on the "live" (active) end of the heating element it will create a short-circuit.
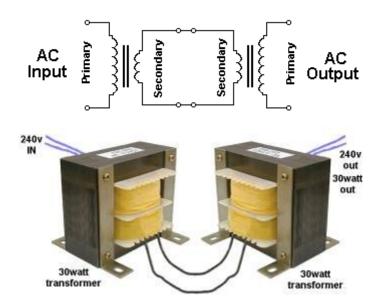
However when the appliance is connected to the mains via an **isolation transformer**, you can touch an earthed soldering iron on either end of the heater as both leads from the isolation transformer are "floating."

Note: As soon as you earth one lead of the output an isolation transformer, the other lead becomes "active."

You can make your own **Isolation Transformer** by connecting two identical transformers "back-to-back."

The following diagram shows how this is done:

You can use any transformers providing the primary and secondary voltages are the same. The current capability of the secondary winding does not matter. However if you want a supply that has almost the same voltage as your "Mains," you need two transformers with the same voltages.

This handy isolation transformer will provide you with "Mains Voltage" but with a limited current.

In other words it will have a limited capability to supply "wattage." If you are using two 15VA transformers, you will only be able to test an appliance rated at 15 watts. This has some advantages and some disadvantages.

If you are working on a project, and a short-circuit occurs, the damage will be limited to 15 watts.

If you are using two transformers with different VA ratings, the lower rating will be the capability of the combination.

If the secondaries are not equal, you will get a higher or lower "Mains Voltage."

If you get two transformers from TVs or Monitors, with a rating on the compliance plate of 45 watts, or 90 watts, you can assume the transformers are capable of delivering this wattage and making an isolation transformer will enable you to test

similar items with the safety of being isolated from the mains.
Colin Mitchell designs a lot of "LED lighting lamps" that are connected directly to the mains. He always works with an isolating transformer, just to be safe. Working on exposed "mains" devices is extremely nerve-wracking and you have to be very careful.
The isolation transformer will prevent a BIG EXPLOSION.

DETERMINING THE SPECS OF A TRANSFORMER
Suppose you have a "mains transformer" with unknown output voltages and unknown current capability.
You must be sure it is a mains transformer designed for operation on 50Hz or 60Hz. Switch-Mode transformers operate at frequencies 40kHz and higher and are not covered in this discussion.
To be on the safe-side, connect the unknown transformer to the output of your isolating transformer.
Since the transformer will take almost no current when not loaded, the output voltages it produces will be fairly accurate. Measure the input AC voltage and output AC voltage.
If the transformer has loaded your isolating transformer it will be faulty.
Mains transformers are approx 15VA for 500gm, 30VA for 1kgm 50VA for 2kgm and and 100VA for 2.5kgm.
VA stands for Volts-Amps and is similar to saying watts. Watts is used for DC circuits, while VA refers to AC circuits.
Once you have the weight of the transformer and the output voltage, you can work out the current capability of the secondary.
For transformers up to 30vA, the output voltage on no-load is 30% higher than the final "loaded voltage."
This is due to the poor regulation of these small devices.
If the transformer is 15VA and the output voltage is 15v AC, the current will be 1 amp AC.
You can check the "quality" of the transformer, (the regulation) by fully loading the output and measuring the final voltage. If the transformer has a number of secondaries, the VA rating must be divided between all the windings.
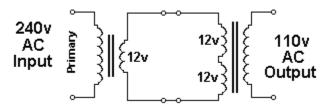
240v to 110v ISOLATION TRANSFORMER
Here's how to create a 110v isolating transformer:
Find a 240v:12v transformer.
Now find a transformer that has two secondary windings, such as 240v:12v+12v.
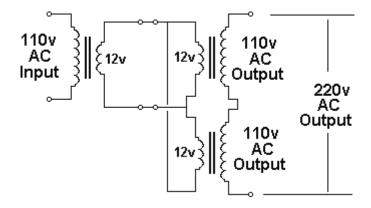
Connect the two transformers as shown in the circuit above. If the output is zero, connect ONE of the 12v windings of the second transformer around the other way.

110v to 240v ISOLATION TRANSFORMER

A 110v to 240v isolation transformer can be created by connecting 3 identical transformers as shown in the diagram above. If the output is zero, connect one of the outputs around the other way.

TRANSFORMER RATINGS

Question from a reader:

I have a 28v - 0 - 28v transformer @3amps. Does this mean each side is 1.5 amps?

The transformer is called CENTRE TAPPED and is shown in figures B and C.

It is designed to be connected to two diodes so each winding takes it in turn to deliver the current as shown in diagram C and the output will be 28v AC at 3 amps. The 28v and 3 amp are both AC values.

If you connect across both outside wires, the output will be 56v at 1.5 amp.

This is because the transformer has a VA rating of 28 x 3 = 84VA. This is very similar to the term "watts."

When the 28v AC is rectified and smoothed, it becomes 28 x 1.4 = 39v (minus 0.6v across the diode) and since the transformer has a rating of 84 VA, the current must be reduced to 84/39 = 2.1 amps to maintain the VA rating.

Some transformers are specified as say: 12v - 0 - 12v, but the wiring diagram is shown as "A." This transformer should be specified as 12v + 12v as the secondaries are separate.
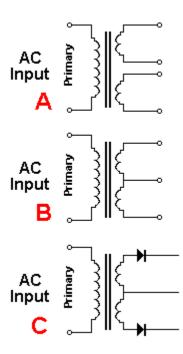
12v - 0 - 12v means the two secondary windings are NOT separate.

It does not make any difference to the output voltage and current, if the windings are separate or joined. The only difference is 12v +12v can be turned into two separate 12v outputs.

If you do not know the output current for a particular transformer, go to the website of electronic parts suppliers and compare the weight of your transformer with others. This will give you a VA rating and you can work out the current, once you know the output voltage.

Note: the output current finishes up ONLY 60% of the rating on the transformer tag because the rating is an **AC RATING**.

With 2 separate secondaries, you can parallel the outputs to get double the current, but don't forget 12v + 12v @ 3amp means 12v in parallel with 12v will provide 2amp DC and the DC voltage will be about 17v.

CURRENT TRANSFORMER

A Current Transformer is really an ordinary transformer.
All transformers produce a CURRENT output and a VOLTAGE output.
If you put an ammeter across the secondary, the current will increase through the meter when the primary voltage is increased.
This is because the output voltage will increase and this voltage will allow a higher current to flow.
WHY DETECT CURRENT? Why not voltage?
Because the voltage of say the "240v AC" is always 240v but the current can increase from say 1 amp to nearly 15 amps, depending what appliance is connected. So it is pointless measuring voltage.
A Current Transformer is a step-up transformer. When we say step-up and step-down, we are referring to the voltage - comparing the primary voltage to the secondary voltage. (Most transformers on the "mains" are step-down transformers and are used as power supplies to laptops, phone chargers etc.) Even a welding transformer is a step-down device and produces about 20v to 70v, while the current can be as high as 100 amps. This current is higher than the mains will deliver and is needed to melt the metal at the point of contact of the probe and the item being welded.
A Current Transformer is a step-up transformer. The primary consists of a single turn (or maybe 2 - 5 turns) and the secondary has 100 turns (or more).
This means the voltage seen by the primary will be increased 100 times and appear as anything from a few hundred millivolts to a few volts, depending on the quality of the coupling. (the magnetic coupling between the wire through the centre of the core, the quality of the core to transfer this magnetic flux to the secondary turns.)
This voltage is then passed to a low value resistor, where the voltage is reduced to a level that suits the detection circuit and the resulting millivolts is interpreted as current in the wire being tested.

Recapping:

The reading on the secondary has no relation to the current in the primary. We need to add a LOAD RESISTOR and create a table before we can use the transformer.
There is no such thing as a CURRENT TRANSFORMER. It is really an INSTRUMENT TRANSFORMER and the scale has been marked in units of CURRENT after measurements have been made.   (INSTRUMENT TRANSFORMER means it is a device that helps us to produce a connection between current flowing through a wire and a reading on a meter or display).
If we connect a load to the secondary, (say an ammeter), it will produce a reading that increases when the current through the single primary turn is increased. That's because the ammeter is a LOAD. But the reading is meaningless until be calibrate the scale.

Now, lets look at the primary.
A wire (or cable) through the centre of the core is counted as one turn. If the turn is wrapped around the core, the coupling will be improved, but if we always use a straight wire, it does not matter where it is positioned inside the centre of the core.
It does not matter if the magnetic interaction of the flux from the wire is good or bad, we just have to keep to the same way of using the transformer.
The calibration can be done with any poor coupling and the result will be accurate for all future readings.
If a low-value resistor is placed across the secondary, the voltage across this resistor will increase and also the current through it will increase. But we are not going to measure the current through the resistor. We are going to measure the voltage across the resistor and by taking lots of reading we will finish up with a scale or table and this is called CALIBRATION. The results will be equated to the current flowing through the primary wire (primary turn).
A clamp meter uses a current transformer and the jaws must be closed completely and cleanly for the flux to flow around the core and produce a reading in the secondary.
Dirt in the jaws will reduce the reading considerably.
You cannot measure the current in a "power cord" because it contains both the active and neutral wires.
Even though the current is a maximum in both conductors at the same time, the current is flowing in two different directions and the magnetic flux produced by one conductor is clockwise and the other is anticlockwise and they are cancelled by each other.
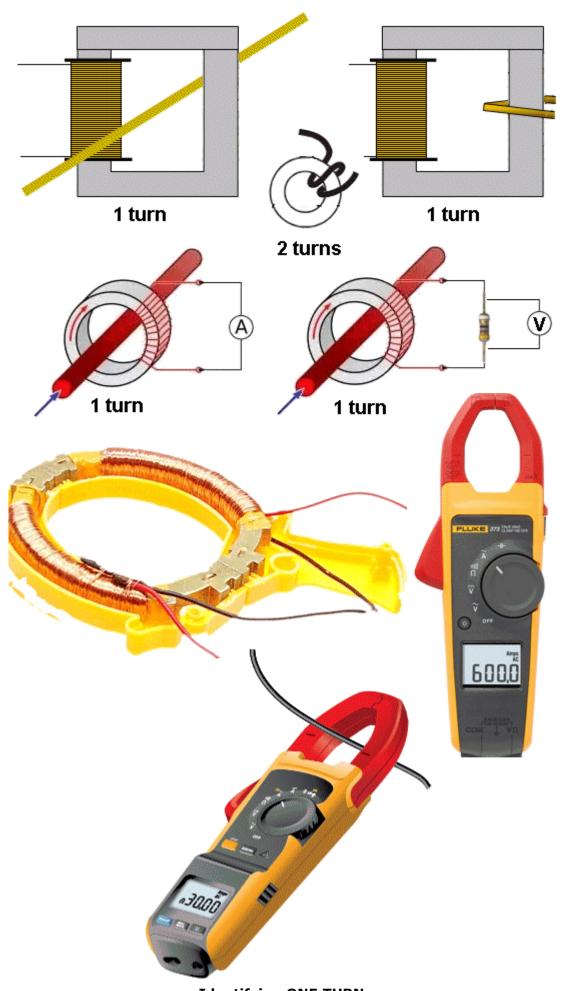
1 turn

2 turns

1 turn

1 turn

1 turn
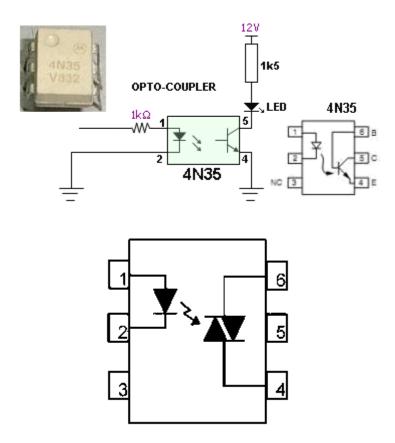
Identifying ONE TURN.
The quality (the coupling) of a single STRAIGHT wire

through the centre of a core is very poor
but if all readings are taken with this amount
of coupling, the readings will be accurate, as the
calibrations have been done with this arrangement.

OPTO ISOLATORS and OPTO COUPLERS

Opto Isolators and Opto Couplers are the same thing. A common opto-coupler is 4N35. It is used to allow two circuits to exchange signals yet remain electrically isolated. The signal is applied to the LED, which shines on a silicon NPN photo-transistor in the IC.
The light is proportional to the signal, so the signal is transferred to the photo transistor to turn it on a proportional amount. Opto-couplers can have Light Activated SCR's, photodiodes, TRIAC's and other semiconductor devices as an output. The 4N35 opto-coupler schematic is shown below:

An opto-Coupler using a TRIAC
Note: the pinout is different to 4N35

TESTING AN OPTO COUPLER

Most multimeters cannot test the LED on the input of an opto-coupler because the ohms range does not have a voltage high enough to activate the LED with at least 2mA.
You need to set-up the test-circuit shown above with a 1k resistor on the input and 1k5 on the output. When the 1k is connected to 12v, the output LED will illuminate. The opto-coupler should be removed from circuit to perform this test.
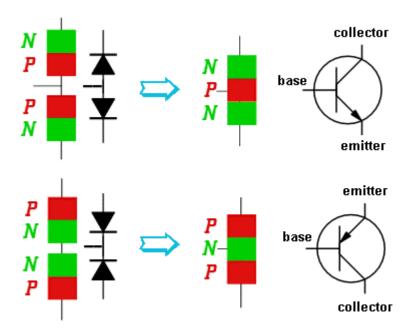
TRANSISTORS

Transistors are solid-state devices and although they operate completely differently to a diode, they appear as two back-to-back diodes when tested.
There are basically 2 types of transistor NPN and PNP.
A transistor is sometimes referred to as BJT (Bi-polar Junction Transistor) to distinguish it from other types of transistor such as Field Effect transistor, Programmable Unijunction Transistor and others.

In the following diagram, two diodes are connected together and although the construction of a transistor is more complex, we see the transistor as two diodes when testing it.

A TRANSISTOR APPEARS AS TWO DIODES WHEN TESTING IT

All transistors have three leads. Base (b), Collector (c), and Emitter (e).
For an NPN transistor, the arrow on the emitter points away from the base.
It is fortunate that the arrow on both symbols points in the direction of the flow of current (Conventional Current) and this makes it easy to describe testing methods using our simplified set of instructions. The symbols have been drawn exactly as they appear on a circuit diagram.

All transistors **are the same** but we talk about digital and analogue transistors.
There is no difference between the two.
The difference is the circuit. And the only other slight difference between transistors is the fact that some have inbuilt diodes and resistors to simplify the rest of the circuit.
All transistors work the same way. The only difference is the amount of amplification they provide, the current and voltage they can withstand and the speed at which they work. For simple testing purposes, they are all the same.

NPN transistors are the most common and for an NPN transistor, the following applies.
(the opposite applies for PNP)
To test a transistor, there is **one thing** you have to know:
When the base voltage is higher than the emitter, current flows though the collector-emitter leads.
As the voltage is increased on the base, nothing happens until the voltage reaches 0.55v. At this point a very small current flows through the collector-emitter leads. As the voltage is increased, the current-flow increases. At about 0.75v, the current-flow is a MAXIMUM. (can be as high as 0.9v). That's how it works. A transistor also needs **current** to flow into the base to perform this amplifying function and this is the one feature that separates an ordinary transistor from a FET.

If the voltage on the base is 0v, then instantly goes to 0.75v, the transistor initially passes NO current, then FULL current. The transistor is said to be working in its two states: OFF then ON (sometimes called: "cut-off" and "saturation"). These are called digital states and the transistor is said to be a **DIGITAL TRANSISTOR** or a **SWITCHING TRANSISTOR** , working in **DIGITAL MODE.**

If the base is delivered 0.5v, then slowly rises to 0.75v and slowly to 0.65v, then

0.7v, then 0.56v etc, the transistor is said to be working in ANALOGUE MODE and the transistor is an **ANALOGUE TRANSISTOR**.
Since a transistor is capable of amplifying a signal, it is said to be an active device. Components such as resistors, capacitors, inductors and diodes are not able to amplify and are therefore known as passive components.

In the following tests, use your finger to provide the **TURN ON** voltage for the base (this is 0.55v to 0.7v) and as you press harder, more current flows into the base and thus more current flows through the collector-emitter terminals. As more current flows, the needle of the multimeter moves UP-SCALE.

TESTING A TRANSISTOR ON A DIGITAL METER
Testing a transistor with a **Digital Meter** must be done on the "DIODE" setting as a digital meter does not deliver a current through the probes on some of the resistance settings and will not produce an accurate reading.
The "DIODE" setting must be used for diodes and transistors. It should also be called a "TRANSISTOR" setting.

TESTING AN unknown TRANSISTOR
The first thing you may want to do is test an unknown transistor for COLLECTOR, BASE AND EMITTER. You also want to perform a test to find out if it is NPN or PNP. That's what this test will provide.
You need a cheap multimeter called an ANALOGUE METER - a multimeter with a scale and pointer (needle).
It will measure resistance values (normally used to test resistors) - (you can also test other components) and Voltage and Current. We use the resistance settings. It may have ranges such as "x10" "x100"   "x1k"   "x10"
Look at the resistance scale on the meter. It will be the top scale.
The scale starts at zero on the right and the high values are on the left. This is opposite to all the other scales.
When the two probes are touched together, the needle swings FULL SCALE and reads "ZERO." Adjust the pot on the side of the meter to make the pointer read exactly zero.

How to read: "x10" "x100" "x1k" "x10"
Up-scale from the zero mark is "1"
When the needle swings to this position on the "x10" setting, the value is 10 ohms.
When the needle swings to "1" on the "x100" setting, the value is 100 ohms.
When the needle swings to "1" on the "x1k" setting, the value is 1,000 ohms = 1k.
When the needle swings to "1" on the "x10k" setting, the value is 10,000 ohms = 10k.
Use this to work out all the other values on the scale.
Resistance values get very close-together (and very inaccurate) at the high end of the scale. [This is just a point to note and does not affect testing a transistor.]

Step 1   - FINDING THE BASE and determining NPN or PNP
Get an unknown transistor and test it with a multimeter set to "x10"
Try the 6 combinations and when you have the black probe on a pin and the red probe touches the other pins and the meter swings nearly full scale, you have an NPN transistor. The black probe is BASE
If the red probe touches a pin and the black probe produces a swing on the other two pins, you have a PNP transistor. The red probe is BASE
If the needle swings FULL SCALE or if it swings for more than 2 readings, the transistor is **FAULTY**.

This is an NPN transistor
The black probe is the BASE

This is a PNP transistor
The red probe is the BASE

Step 2   - FINDING THE COLLECTOR and EMITTER
Set the meter to "x10k."
For an NPN transistor, place the leads on the transistor and when you press hard on the two leads shown in the diagram below, the needle will swing almost full scale.

Set to "x10k"

This is the
POSITIVE
of the battery
in the meter

push
HERE

This is an
NPN
transistor

This is the NEGATIVE
of the battery in the meter

Here is the equivalent circuit:

+

The finger is
"turning ON
the transistor"

collector

base

emitter

You have found the Collector Base and Emitter

For a PNP transistor, set the meter to "x10k" place the leads on the transistor and when you press hard on the two leads shown in the diagram below, the needle will swing almost full scale.

Set to "x10k"

This is the
POSITIVE
of the battery
in the meter

This is a
PNP
transistor

This is the NEGATIVE
of the battery in the meter

push
HERE

Here is the equivalent circuit:

The finger is
"turning ON
the transistor"

emitter

base

collector

You have found the Collector Base and Emitter

SIMPLEST TRANSISTOR TESTER
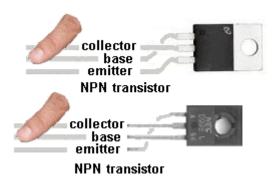
The simplest transistor tester uses a 9v battery, 1k resistor and a LED (any colour).
Keep trying a transistor in all different combinations until you get one of the circuits
below. When you push on the two leads, the LED will get brighter.
The transistor will be NPN or PNP and the leads will be identified:

1k

LED

k

push
HERE

base

collector

emitter

This is an
NPN
transistor

Here is the equivalent circuit:

9v

The finger is
"turning ON
the transistor"

1k

LED
k
c

b

9v

e

NPN

1k

LED

This is a
PNP
transistor

k

emitter

base

collector

push
HERE

Here is the equivalent circuit:

9v

PNP

The finger is
"turning ON
the transistor"

1k

LED
e k

b

9v

c

The leads of some transistors will need to be bent so the pins are in the same positions as shown in the diagrams. This helps you see how the transistor is being turned on. This works with NPN, PNP transistors and Darlington transistors.

collector
base
emitter
NPN transistor

collector
base
emitter
NPN transistor

HEATSINKING

Heat generated by current flowing between the collector and emitter leads of a transistor causes its temperature to rise. This heat must be conducted away from the transistor otherwise the rise may be high enough to damage the P-N junctions inside the device. Power transistors produce a lot of heat, and are therefore usually mounted on a piece of aluminium with fins, called a **HEATSINK**.
This draws heat away, allowing it to handle more current. Low-power signal transistors do not normally require heat sinking. Some transistors have a metal body or fin to connect to a larger heatsink. If the transistor is connected to a heatsink with a mica sheet (mica washer), it can be damaged or cracked and create a short-circuit. (See Testing Mica Washers). Or a small piece of metal may be puncturing the mica. Sometimes white compound called **Heatsink Compound** is used to conduct heat through the mica. This is very important as mica is a very poor conductor of heat and the compound is needed to provide maximum thermal conduction.

TRANSISTOR FAILURE

Transistor can fail in a number of ways. They have forward and reverse voltage ratings and once these are exceeded, the transistor will ZENER or conduct and may fail. In some cases a high voltage will "puncture" the transistor and it will fail instantly. In fact it will fail much faster via a voltage-spike than a current overload.

It may fail with a "short" between any leads, with a collector-emitter short being the most common. However failures will also create shorts between all three leads.
A shorted transistor will allow a large current to flow, and cause other components to heat up.
Transistors can also develop an open circuit between base and collector, base and emitter or collector and emitter.
The first step in identifying a faulty transistor is to check for signs of overheating. It may appear to be burnt, melted or exploded. When the equipment is switched off, you can touch the transistor to see if it feels unusually hot. The amount of heat you feel should be proportional to the size of the transistor's heat sink. If the transistor has no heat sink, yet is very hot, you can suspect a problem.
DO NOT TOUCH A TRANSISTOR IF IT IS PART OF A CIRCUIT THAT CARRIES 240VAC. Always switch off the equipment before touching any components.

TRANSISTOR REPLACEMENT

If you can't get an exact replacement, refer to a transistor substitution guide to identify a near equivalent.

The important parameters are:
- Voltage
- Current
- Wattage
- Maximum frequency of operation

The replacement part should have parameters equal to or higher than the original.

Points to remember:
- Polarity of the transistor i.e. PNP or NPN.
- At least the same voltage, current and wattage rating.
- Low frequency or high frequency type.
- Check the pinout of the replacement part
- Use a desoldering pump to remove the transistor to prevent damage to the
    printed circuit board.
- Fit the heat sink.
- Check the mica washer and use heat-sink compound
- Tighten the nut/bolt - not too tight or too loose.
- Horizontal output transistors with an integrated diode should be replaced with the
    same type.

DIGITAL TRANSISTORS

There is no such thing as a DIGITAL TRANSISTOR or an AUDIO TRANSISTOR.
All transistors are just "TRANSISTORS" and the surrounding components as well as the type of signal, make the transistor operate in DIGITAL MODE or ANALOGUE MODE.
But we have some transistors that have inbuilt resistors to make them suitable for connecting to a digital circuit without the need for a base resistor.
Here is the datasheet for an NPN transistor BCR135w and PNP datasheet for BCR185w.
These transistors are called "Digital Transistors" because the "base lead" can be connected directly to the output of a digital stage. This "lead" or "pin" is not really the base of the transistor but a 4k7 (or 10k) resistor connected to the base allows the transistor to be connected to the rest of a digital circuit.
You cannot actually get to the base. The resistor(s) are built into the chip and the transistor is converted into a "Digital Transistor" because it will accept 5v on the "b"
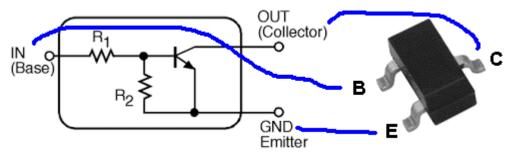
lead.
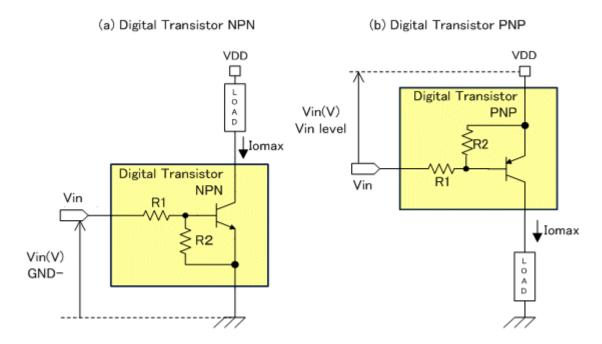
The 47k is not really needed but it makes sure the transistor is fully turned OFF if the signal on the "b" lead is removed (in other words - if the input signal is converted to a high-impedance signal - see tri-state output from microcontrollers for a full explanation).

This transistor is designed to be placed in a circuit where the input changes from low to high and high to low and does not stop mid-way. This is called a DIGITAL SIGNAL and that is one reason why the transistor is called a digital transistor. (However you could stop half-way but the transistor may heat up and get too hot).

Any transistor placed in a digital circuit can be called a "digital transistor" but it is better to say it is operating in DIGITAL MODE.
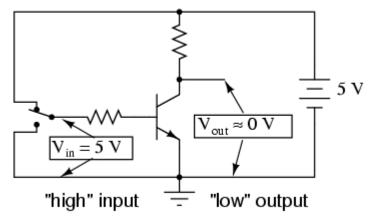
The digital transistor has two resistors included inside the case
R1 is about 10k and R2 is approx 47k

Equivalent circuit of the digital transistor

These transistors can be made to work in analogue circuits because they are ordinary transistors with a 10k base resistor, but you will have to know what you are doing.

Transistor in saturation

The circuit above shows the digital transistor is designed to allow a voltage of 5v to be supplied to the "base" pin and the transistor will Fully Conduct.
This type of transistor saves putting a base resistor on the PC board.
It can be tested just like a normal transistor but the resistance between base and emitter will be about 5k to 50k in both directions. If the collector-emitter is low in both directions the transistor is damaged.
Here's how to look at how the transistor works:
The 10k resistor on the base will allow 0.5mA to flow into the base. But the 47k will reduce this to 0.4mA. If the transistor has a gain of 100, the collector-emitter current can be 40mA.
To determine the current capability of the transistor, connect 100R load and turn the transistor ON. This will allow about 100mA for the collector-emitter current. Measure the collector-emitter voltage. If it is more than 0.5v, the transistor is OVER-LOADED.
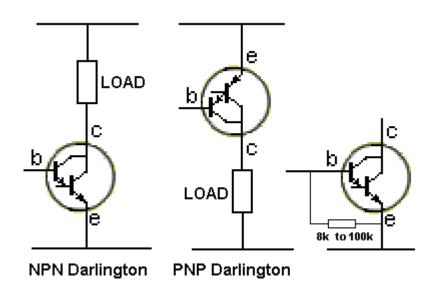
DARLINGTON TRANSISTORS

A DARLINGTON TRANSISTOR is two transistors in a single package with three leads. They are internally connected in cascade so the gain of the pair is very high. This allows a very small input signal to produce a large signal at the output. They have three leads (Base, Collector and Emitter and can be PNP or NPN) and are equivalent to the leads of a standard individual transistor, but with a very high gain. The second advantage of a Darlington Transistor is its high input impedance. It puts very little load on the previous circuit.
Some Darlington transistors have a built-in diode and/or built-in resistor and this will produce a low reading in both directions between the base and emitter leads.

NPN Darlington   PNP Darlington

Darlington transistors are tested the same as an ordinary transistor and a multimeter will produce about the same deflection, even though you will be measuring across two junctions, (and a base-emitter resistor is present).

HORIZONTAL OUTPUT TRANSISTORS, SWITCH-MODE TRANSISTORS, FLYBACK TRANSISTORS, POWER TRANSISTORS, VERTICAL TRANSISTORS . . . .

These are all names given to a transistor when it is used in a particular circuit. ALL these transistors are the same for testing purposes.
We are not testing for gain, maximum voltage, speed of operation or any special feature. We are just testing to see if the transistor is completely faulty and SHORTED.
A transistor can have lots of other faults and the circuit **using the transistor** is the best piece of TEST EQUIPMENT as it is detecting the fault.
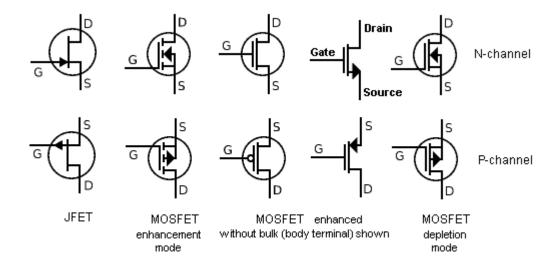
TESTING MOSFETs and FETs

MOSFETs and **JFETs** are all part of the **FET** family.
MOSFET stands for **Metal Oxide Semiconductor Field Effect Transistor.**
FETs operate exactly the same as a "normal" transistor except they have different names for the input and output leads and the voltage between the gate and the source has to between 2v to 5v for the device to turn on fully. A FET requires almost NO CURRENT into the Gate for it to turn on and when it does, the voltage between drain and source is very low (only a few mV). This allows them to pass very high currents without getting hot. There is a point where they start to turn on and the input voltage must rise higher than this so the FET turns on FULLY and does not get hot.
Field Effect Transistors are difficult to test with a multimeter, but "fortunately" when a power **MOSFET** blows, it is completely damaged. All the leads will show a short circuit. 99% of bad **MOSFETs** will have GS, GD and DS shorted.
The following symbols show some of the different types of MOSFETs:
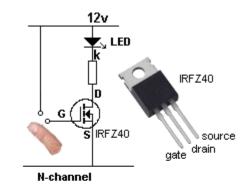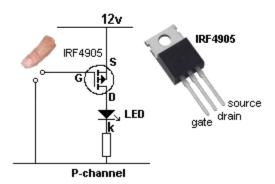
| JFET | MOSFET enhancement mode | MOSFET enhanced without bulk (body terminal) shown | MOSFET depletion mode |

Most **MOSFET** transistors cannot be tested with a multimeter. This due to the fact that the Gate needs 2v - 5v to turn on the device and this voltage is not present on the probes of either meter set to any of the ohms ranges.
You need to build the following Test Circuit:

N-channel

P-channel

Touching the Gate will increase the voltage on the Gate and the MOSFET will turn ON and illuminate the LED. Removing your finger will turn the LED off.

Large devices such as the TO-220 types shown above do not like static electricity on the gate and you have to be careful not to "spike" the gate with any static. Generally this type of device is not "super sensitive" and you can use your finger or a large value resistor.
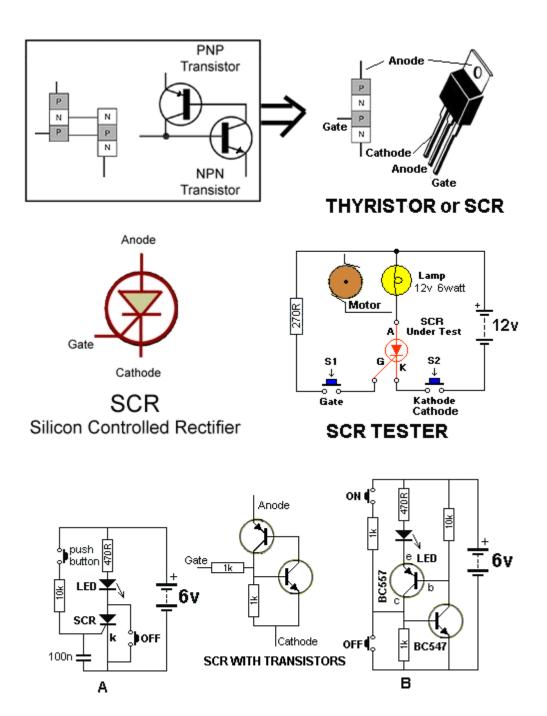
When replacing one of these devices, there are 2 things to match-up.

Voltage and Current.

In most cases, the "turn-ON" resistance (the resistance between Source and Drain) will be the same (something like 22 milli ohms) and the speed of operation will be ok.

Check the voltage needed to turn the gate ON and make sure you can supply the required voltage.

SILICON CONTROLLED RECTIFIERs (SCR)

The **Silicon Controlled Rectifier** (SCR) is a semiconductor device that is a member of a family of control devices known as **Thyristors**. It is a 3-leaded device and when a small current enters the Gate, the **thyristor** turns on. AND STAYS ON. It only conducts current between Anode and Cathode in one direction and it is mainly only used in DC circuits. When it is used with AC, it will only conduct for a maximum of half the cycle.

PNP Transistor

NPN Transistor

Anode

Gate

Cathode

Anode

Gate

THYRISTOR or SCR

Anode

Gate

Cathode

SCR

Silicon Controlled Rectifier

Motor

Lamp 12v 6watt

SCR Under Test

12v

S1

Gate

G

K

S2

Kathode Cathode

SCR TESTER

push button

LED

SCR

OFF

100n

6v

A

Anode

Gate

Cathode

SCR WITH TRANSISTORS

ON

LED

BC557

OFF

BC547

6v

B

To understand how an SCR "latches" when the gate is provided with a small current, we can replace it with two transistors as shown in diagram B above. When the ON button is pressed, the BC547 transistor turns on. This turns ON the BC557 and it takes over from the action of the switch.
To turn the circuit off, the OFF button removes the voltage from the base of the BC547.

Testing an SCR

An **SCR** can be tested with some multimeters but a minimum current Anode-to-Cathode is needed to keep the device turned on. Some multimeters do not provide this amount of current and the **SCR Tester** circuit above is the best way to test these devices.
Shorted SCRs can usually be detected with an ohmmeter check (SCRs usually fail shorted rather than open).
Measure the anode-to-cathode resistance in both the forward and reverse direction; a good SCR should measure near infinity in both directions.
Small and medium-size SCRs can also be gated ON with an ohmmeter (on a digital meter use the Diode Check Function). Forward bias the SCR with the ohmmeter by connecting the black (-) lead to the anode and the red (+) lead to the cathode

(because the + of the battery is connected to the negative lead, in most analogue multimeters). Momentarily touch the gate lead to the anode while the probes are still touching both leads; this will provide a small positive turn-on voltage to the gate and the cathode-to-anode resistance reading will drop to a low value. Even after removing the gate voltage, the SCR will stay conducting. Disconnecting the meter leads from the anode or cathode will cause the SCR to revert to its non-conducting state.

When making the above test, the meter impedance acts as the SCR load. On larger SCRs, it may not latch ON because the test current is not above the SCR holding current.

Using the SCR Tester
Connect an SCR and press Switch2. The lamp should not illuminate. If it illuminates, the SCR is around the wrong way or it is faulty.

Keep Switch 2 PRESSED. Press Sw1 very briefly. The lamp or motor will turn ON and remain ON. Release Sw 2 and press it again. The Lamp or motor will be OFF.
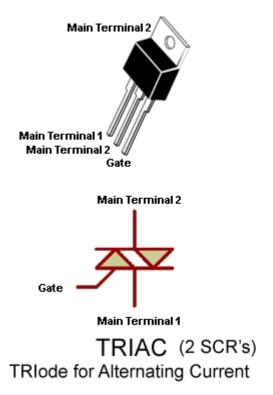
TRIACs

A triac is a bidirectional, three-terminal dual, back-to-back thyristor (SCR) switch. This device will conduct current in both directions when a small current is constantly applied to the Gate.

If the gate is given a small, brief, current during any instant of a cycle, it will remain triggered during the completion of the cycle until the current though the Main Terminals drops to zero.

This means it will conduct both the positive and negative half-cycles of an AC waveform. If it is tuned on (with a brief pulse) half-way up the positive waveform, it will remain on until the wave rises and finally reaches zero. If it is then turned on (with a brief pulse) part-way on the negative wave, the result will be pulses of energy and the end result will be about 50% of the full-energy delivered at a rate of 100 times per second for a 50HZ supply.
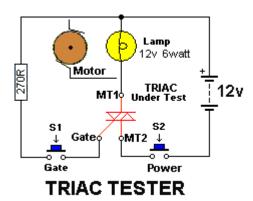
TRIACs are particularly suited for AC power control applications such as motor speed control, light dimmers, temperature control and many others.

TRIAC (2 SCR's)
TRIode for Alternating Current

Using the TRIAC Tester
Connect a TRIAC and press Switch2. The lamp should not illuminate. If it illuminates, the TRIAC is faulty.
Keep Switch 2 PRESSED. Press Sw1 very briefly. The lamp or motor will turn ON and remain ON. If the lamp does not turn on, reverse the TRIAC as the current into the gate must produce a slight voltage between **Gate** and **Main Terminal 1**.
Release Sw 2 and press it again. The Lamp or motor will be OFF.

TRIAC TESTER

MICA WASHERS AND INSULATORS
Plastic insulating sheets (washers) between a transistor and heatsink are most often made from mica but some are plastic and these get damaged over a period of time, turn dark and become cracked.
The plastic eventually becomes carbonized and conducts current and can affect the operation of the appliance. You can see the difference between a mica sheet (washer) and plastic by looking where it extends from under the transistor. Replace all plastic insulators as they eventually fail.

SPARK GAPS
Some TV's and monitors with a CRT (picture tube), have spark gaps either on the socket at the end of the tube or on the chassis.
These can consist of two wires inside a plastic holder or a glass tube or special resistive device.
The purpose of a spark gap is to take any flash-over (from inside the tube), to earth. This prevents damage to the rest of the circuit.
However if the tube constantly flashes over, a carbon track builds up between the wires and effectively reduces the screen voltage. This can cause brightness and/or focus problems. Removing the spark-gap will restore the voltage.
These are not available as a spare component and it's best to get one from a discarded chassis.

CO-AX CABLES
Co-Ax cables can produce very high losses and it seems impossible that a few metres of cable will reduce the signal. The author has had a 3 metre cable reduce the signal to "snow" so be aware that this can occur. Faults can also come from a splitter and/or balun as well as dirty plugs and sockets. This can result in very loud bangs in the sound on digital reception.

TESTING EARTH LEAKAGE DETECTORS or
Residual Current Devices or
Ground Fault Circuit Interrupters or GFCI
An Earth Leakage Detector or Sensor is a circuit designed to continuously monitor the imbalance in the current in a pair of load carrying conductors.
These two conductors are normally the Active and Neutral. Should the imbalance current reach 30mA the sensor will "trip" and remove the voltage (and current) from

the line being monitored.
Some detectors will trip at 15mA.
You cannot alter the sensitivity of the device however there are a number of faults in these devices that can be fixed.
In some devices the contact pressure for the 10Amp or 15 Amp contacts is very weak and they arc and produce an open circuit. The result is this: When you press the rest button, power is not restored to the output.
Clean the contacts with a small file and bend the metal strips to the contacts so they make a very strong contact.
The other fault is the trip mechanism.
The magnetism from the coil does not allow the pin to move and "trip" the contacts. It may be due to a small metal filing or the pin not moving freely enough.
All good Earth Leakage Detectors have a TEST BUTTON. This connects a resistor between the active line and earth so that 15mA or 30mA flows.
The detector should trip immediately. Make sure the trigger mechanism trips when the test button is pressed.
None of the electronics in the detector can be replaced however you can test the mechanical operation and the pressure on the contacts when the unit is removed from the power. Do not work on the device when it is connected to the mains.

TESTING CELLS AND BATTERIES

There is an enormous number of batteries and cells on the market and a number of "battery testers." Instead of buying a battery tester that may give you a false reading, here is a method of testing cells that is guaranteed to work.
There are two types of cell: a **rechargeable** cell and a non rechargeable cell.

The easiest way to test a **rechargeable** cell is to put a group of them in an appliance and use them until the appliance "runs down" or fails to work. If you consider the cells did not last very long, remove them and check the voltage of each cell. The cell or cells with the lowest voltage will be faulty. You can replace them with new cells or good cells you have in reserve.
There is no other simple way to test a rechargeable cell.
You cannot test the "current of a cell" by using an ammeter. A rechargeable cell can deliver 10 amps or more, even when nearly discharged and you cannot determine a good cell for a faulty cell.

Dry cells are classified as "non-rechargeable" cells.
DRY CELLS and MANGANESE CELLS are the same thing. These produce 1.5v per cell (manganese means the Manganese Dioxide depolariser inside the cell. All "dry cells" use manganese dioxide).
ALKALINE CELLS produce between 2 - 10 times more energy than a "dry cell" and produce 1.5v per cell.
Alkaline cells can fail for no reason at any stage in their life and are not recommended for emergency situations.
The output voltage of some Alkaline cells can fall to 0.7v or 0.9v for not apparent reason.
There are lots of other cells including "button cells," hearing-aid cells, air cells, and they produce from 1.2v to 3v per cell.

Note:
Lithium cells are also called "button cells" and they produce 3v per cell.
Lithium cells are non-rechargeable (they are generally called "button cells") but some Lithium cells can be recharged. These are Lithium-ion cells and generally have a voltage of 3.6v. Some Lithium-ion cells look exactly like 3v Lithium cells, so you have to read the data on the cell before charging.

You cannot test the voltage of a cell and come to any conclusion as to the age of the cell or how much energy remains. The voltage of a cell is characteristic to the chemicals used and the actual voltage does not tell you its condition.
Some "dry cells" deliver 1.5v up to the end of their life whereas others drop to about 1.1v very quickly.

Once you know the name of the cell that drops to 1.1v, avoid them as the operation of the equipment "drops off" very quickly.

However if you have a number of different cells and need to know which ones to keep, here's the solution:
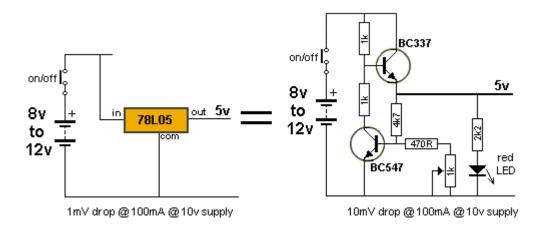
1. Check the voltage and use those with a voltage above 1.1v
2. Next, select 500mA or 10A range on a meter and place the probes on a cell. For a AAA or AA cell, the current should be over 500mA and the needle will swing full scale very quickly.
Keep the testing short as you are short-circuiting the cell but it is the only way to determine the internal impedance of the cell and this has a lot to do with its stage-of-charge.
This will give you a cell with a good terminal voltage and a good current capability.

This also applies to button cells, but the maximum current they will deliver will be less.
If you want to get the last of the energy out of a group of cells they can be used in the following circuits:

1mV drop @100mA @10v supply   10mV drop @100mA @10v supply

TESTING PIEZO DIAPHRAGMS and PIEZO BUZZERS
There are two types of piezo devices that produce a sound.
They are called **PIEZO DIAPHRAGMS** and **PIEZO BUZZERS**.
A **piezo diaphragm** consists of two metal plates with a ceramic material between. The ceramic expands and contracts when an alternating voltage is placed on the two plates and this causes the main plate to "dish" and "bow."
This creates a high-pitched sound. There are no other components inside the case and it requires an AC voltage of the appropriate frequency to produce a sound.
A **piezo buzzer** has a transistor and coil enclosed and when supplied with a DC voltage, the buzzer produces a sound.
Both devices can look exactly the same and the only way to tell them apart is by connecting a 9v battery. One device may have "+' and "-" on the case to indicate it is a piezo buzzer, but supplying 9v will make the buzzer produce a sound while the piezo diaphragm will only produce a "click."

PIEZO DIAPHRAGM   PIEZO BUZZER
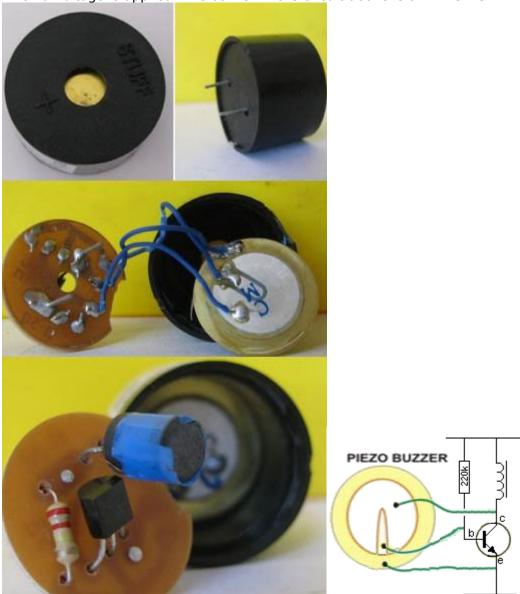
A piezo diaphragm will produce a click
when connected to 9v DC.
A piezo buzzer will produce a tone when
connected to a DC voltage.

How a PIEZO BUZZER WORKS

A Piezo Buzzer contains a transistor, coil, and piezo diaphragm and produces sound when a voltage is applied. The buzzer in the circuit above is a PIEZO BUZZER.

The circuit starts by the base receiving a small current from the 220k resistor. This produces a small magnetic flux in the inductor and after a very short period of time the current does not increase. This causes the magnetic flux to collapse and produce a voltage in the opposite direction that is higher than the applied voltage.

3 wires are soldered to pieces of metal on the top and bottom sides of a ceramic substrate that expands sideways when it sees a voltage. The voltage on the top surface is passed to the small electrode and this positive voltage is passed to the base to turn the transistor ON again. This time it is turned ON more and eventually the transistor is fully turned ON and the current through the inductor is not an INCREASING CURRENT by a STATIONARY CURRENT and once again the magnetic flux collapses and produces a very high voltage in the opposite direction. This voltage is passed to the piezo diaphragm and causes the electrode to "Dish" and produce the characteristic sound. At the same time a small amount is "picked-off" and sent to the transistor to create the next cycle.

TESTING A SPEAKER

A speaker (also called a **loud speaker**) has coil of wire wrapped around a magnet but it does not

touch the magnet as it is wound on a thin cardboard former so that the coil will be pulled closer to the magnet when a current flows in the coil.

When the current flows in the other direction, the coil moves away from the magnet.

The coil is called voice coil and it is connected to a sheet of thin card called a CONE and as the cone vibrates, the speaker reproduces music or noise.

Use a multimeter on a low ohms scale to read the value of resistance of the coil.

It can be as low as 2 ohms or as high as 100 ohms.

Most speakers have an 8R voice coil and the actual resistance may be slightly lower than this.

Some speakers have a resistance of 16R, 32R or 50R and even 75 ohms.

You would think putting a 16R speaker in place of 8R would reduce the sound output, but this is not always the case.

You can even use 50R or 75R and get the same performance.

This may sound amazing, but here is the reason.

The cone is deflected a certain amount due to the current flowing and the number of turns.

These two values are multiplied together to produce a value called AMP-TURNS.

If we have an 8R speaker with 80 turns and 100mA, the result is 0.1 x 80 = 8.

If we use a 16R speaker, the average current flow will be 50mA and the number of turns will be about 160.   The multiplication of 0.05 x 160 = 8.

The author then tried a 50R speaker and the sound output was equal to 8R and the same with 75R speaker.

This might not apply in all situations, but the 75R speaker was slightly larger and the ticking sound form the **Metal Detector** kit was louder than using an 8R mini speaker.

To see if the cone of a speaker is undamaged, push it slightly and it will move towards the magnet. If it does not move, it is bent or damaged. If the cone is scratchy when pushed, it is rubbing against the magnet.

A cone should be able to be pushed and pulled from its rest-state. If not, it will produce a distorted sound.

TESTING A CIRCUIT

Whenever you test a circuit, the TEST EQUIPMENT puts "a load" or "a change" on it. It does not matter if the test equipment is a multimeter, Logic Probe, CRO, Tone Injector or simply a LED and resistor.

There are two things you need to know.

1. The IMPEDANCE of the circuit at the location you are testing, and
2. The amount of load you are adding to the circuit via the test equipment.

There is also one other hidden factor. The test equipment may be injecting "hum" due to its leads or the effect of your body at absorbing hum from the surroundings or the test equipment may be connected to the mains.

These will affect the reading on the test equipment and also any output of the circuit. Sometimes the test equipment will prevent the circuit from working and sometimes it will just change the operating conditions slightly. You have to be aware of this.

The last section of this eBook covers High and Low Impedance and understanding impedance is something you need to know.

The point to note here is the fact that the equipment (and the reading) can be upset by hum and resistance/capacitance effects of test equipment. This is particularly critical in high impedance and high frequency circuits.

TESTING INTEGRATED CIRCUITS   (IC's)

Integrated Circuits can be tested with a LOGIC PROBE. A Logic Probe will tell you if a line is HIGH, LOW or PULSING.
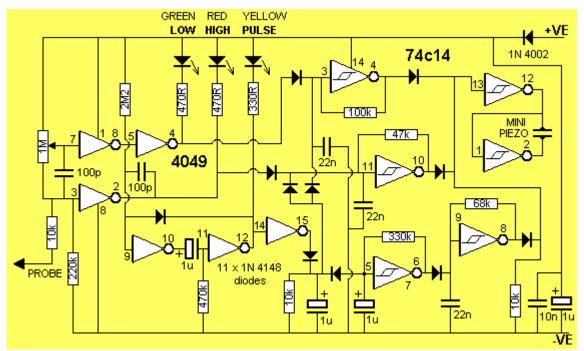
Most logic circuits operate on 5v and a Logic Probe is connected to the 5v supply so the readings are accurate for the voltages being tested.

A Logic Probe can also be connected to a 12v CMOS circuit.

You can make your own Logic Probe and learn how to use it from the following link:
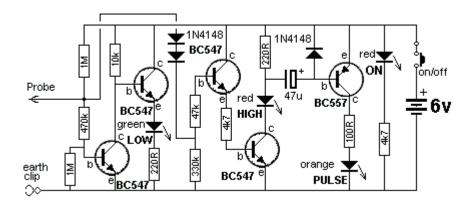
http://www.talkingelectronics.com/projects/LogicProbeMkIIB/LogicProbeMk-IIB.html
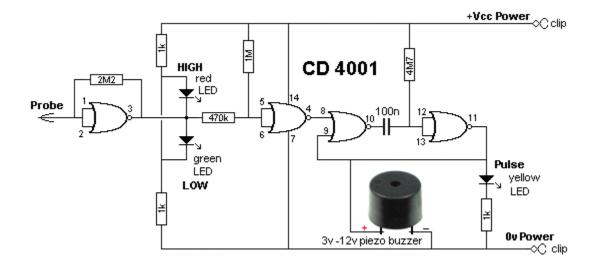
LOGIC PROBE with PULSE
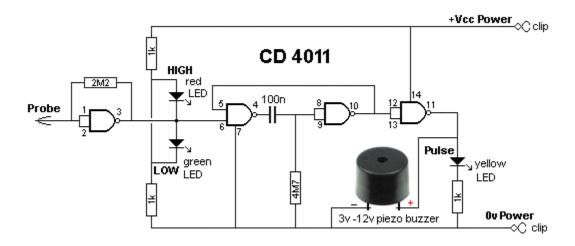
This is a very simple transistor circuit to provide HIGH-LOW-PULSE indication for digital circuits. It can be built for less than $5.00 on a piece of matrix board or on a small strip of copper clad board if you are using surface mount components. The probe will detect a HIGH at 3v and thus the project can be used for 3v, 5v and CMOS circuits.
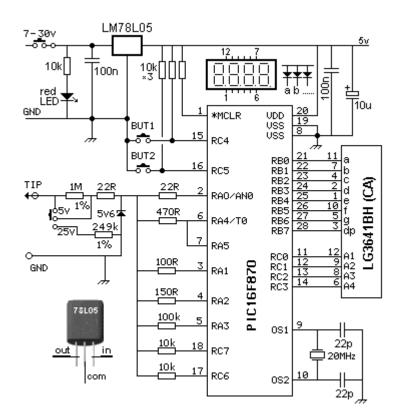
LOGIC PROBE using CD4001 and CD4011

Here is a simple Logic Probe using a single chip. The circuits have been designed for the **CD4001** CMOS quad NOR gate and **CD4011** CMOS NAND gate. The output has an active buzzer that produces a beep when the pulse LED illuminates (the buzzer is not a piezo-diaphragm but an active buzzer containing components).

CD 4001

+Vcc Power ⊸C clip

1k HIGH red LED 1M 4M7

2M2

Probe 470k

green LED LOW

5 14 4 8 10 100n 12 11
6 7 9 13

1k

3v -12v piezo buzzer

Pulse yellow LED

1k

0v Power ⊸C clip

CD 4011

+Vcc Power ⊸C clip

1k HIGH red LED

2M2

Probe

1 3 5 4 100n 8 10 12 14 11
2 6 7 9 13

green LED LOW LED

1k 4M7

3v -12v piezo buzzer

Pulse yellow LED

1k

0v Power ⊸C clip
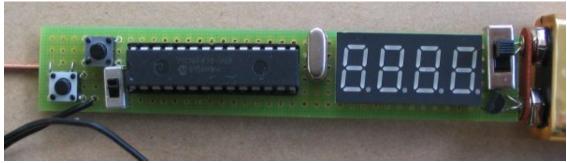
[SUPER PROBE MkII](#) has 20 different features including a Logic Probe, capacitance tester, Inductance tester, and more.

7–30v LM78L05 5v

10k 100n 10k ×3 12   7

red LED a b 100n 10u

GND

BUT1 1 *MCLR   VDD 20
              VSS 19
15 RC4   VSS 8

BUT2
16 RC5

TIP 1M 22R 22R 2 RA0/AN0   RB0 21 11 a
05v 1% 5v6 470R 6 RA4/T0   RB1 22 7 b
25v 249k   RB2 23 4 c
     1%   7 RA5   RB3 24 2 d
GND   RB4 25 1 e
                100R 3 RA1   RB5 26 10 f
                                 RB6 27 5 g
                150R 4 RA2   RB7 28 3 dp
78L05   100k 5 RA3
                                 RC0 11 12 A1
out in   10k 18 RC7   RC1 12 9 A2
                                 RC2 13 8 A3
                10k 17 RC6   RC3 14 6 A4
com
                                 OS1 9 22p
                                        20MHz
                PIC16F870   OS2 10 22p

LG3641BH (CA)

SUPER PROBE MkII Circuit

To test an IC, you need a circuit diagram with waveforms. These diagrams will show the signals and are very handy if a CRO (cathode ray Oscilloscope) is used to diagnose the problem. The CRO will reproduce the waveform and prove the circuit is functioning correctly.
A Logic Probe will just show activity and if an output is not producing a "pulse" or "activity," you should check the power to the IC and test the input line.
It is beyond the scope of this eBook to explain how to diagnose waveforms, however it is important to know if signals are entering and exiting an IC and a Logic Probe is designed for this.

SIGNAL INJECTOR
This circuit is rich in harmonics and is ideal for testing amplifier circuits. To find a fault in an amplifier, connect the earth clip to the 0v rail and move through each stage, starting at the speaker. An increase in volume should be heard at each preceding stage. This Injector will also go through the IF stages of radios and FM sound sections in TV's.

TESTING AUDIO AMPLIFIERS and AUDIO IC's

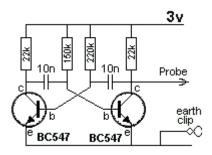The **Super Probe MII** described above has a "noise" function and a tone function that allows you to inject a signal into an audio stage, amplifier (made from discrete components) or an audio chip, and detect the output on a speaker.
Audio stages are very difficult to work-with if you don't have a TONE GENERATOR or SIGNAL INJECTOR.
The signals are very small and not detected by a multimeter.
You can start anywhere in an amplifier and when a tone is heard, you can keep probing until the signal is not present or louder. From this you can work out which way the signal is travelling.
A Signal Injector is very handy for finding shorts and broken wires in switches, plugs, sockets and especially leads to headphones.
You can determine the gain of a stage (amplification) by probing before and after a chip or transistor and listen for the relative increase in volume from the speaker.
You can also use your finger to produce "hum" or "buzz" if a **Signal Injector** is not available.
Nearly all audio problems are plugs, sockets and cracks in the PC board, but finding them takes a lot of time and skill.

TESTING IC's - also called "CHIPS"

An Integrated Circuit is also called a "chip." It might have 8 pins or as many as 40.
Some chips are ANALOGUE. This means the input signal is rising and falling slowly
and the output produces a larger version of the input.
Other chips are classified as DIGITAL and the input starts at 0v and rises to rail
voltage very quickly. The output does exactly the same - it rises and falls very
quickly.
You might think the chip performs no function, because the input and output voltage
has the same value, but you will find the chip may have more than one output and
the others only go high after a number of clock-pulses on the input, or the chip may
be outputting when a combination of inputs is recognised or the output may go HIGH
after a number of clock pulses.

ANALOGUE CHIPS (also see above)
Analogue chips are AUDIO chips or AMPLIFIER chips.
To test these chips you will need three pieces of test equipment:
1. A multimeter - this can be digital or analogue.
2. A Signal Injector
3. A Mini Bench Amplifier.
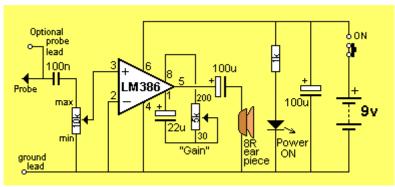
The Mini Bench Amplifier is available as a kit.

MINI BENCH AMPLIFIER

MINI BENCH AMPLIFIER CIRCUIT

Start by locating the power pin with a multimeter.
If the chip is receiving a voltage, you can use the Mini Bench Amplifier to detect an
output.
Connect the Ground Lead of the Mini Bench Amplifier to 0v and touch the Probe tip
on each of the pins.
You will hear faint audio on the Input pin and very loud audio on the Output pin.
If no input is detected, you can use a Signal Injector to produce a tone.

Connect the clip of the Signal Injector to 0v and the probe to the input pin of the amplifier chip. At the same time, connect the Mini Bench Amplifier to the output pin and you will hear a very loud tone.

These pieces of test equipment can also be used to diagnose an amplifier circuit constructed with individual components.

Amplifier circuits using discrete components are very hard to trouble-shoot and these pieces of test equipment make it very easy.

DIGITAL CHIPS

It is always best to have data on the chip you are testing, but if this is not available, you will need three pieces of equipment:

1. A multimeter - this can be digital or analogue.
2. A Logic Probe,
3. A logic Pulser.

Firstly test the chip to see if power is being delivered. This might be anything from 3v3 to 15v.

Place the negative lead of the multimeter on the earth rail of the project - this might be the chassis, or the track around the edge of the board or some point that is obviously 0v.

Try all the pins of the chip and if you get a reading, the chip will have "supply."

Identify pin 1 of the chip by looking for the "cut-out" at the end of the chip and you may find a small dimple below the cut-out (or notch). This is pin 1 and the "power pin" can be directly above or any of the other pins.

Next you need to now if a signal is entering the chip.

For this you will need a LOGIC PROBE.

A Logic Probe is connected to the same voltage as the chip, so it will detect a HIGH and illuminate a red LED.

Connect the Logic Probe and touch the tip of the probe on each pin.

You will not know if a signal is an input or output, however if you get two or more active pins, you can assume one is input and the other is output. If none of the pins are active, you can assume the signal is not reaching this IC.

If only one pin is active, you can assume the chip is called a CLOCK (or Clock Generator). This type of chip produces pulses. If more than two pins are active, you can assume the chip is performing its function and unless you can monitor all the pins at the same time, you don't know what is happening.

This is about all you can do without any data on the chip.

If you have data on the chip, you can identify the input(s) and output(s).

A Logic Probe on each of these pins will identify activity.

A Logic Probe has 3 LEDs. Red LED indicates a HIGH, Green indicates a LOW and Orange indicates a PULSE (activity).

Some Logic Probes include a piezo and you can hear what is happening, so you don't take your eyes off the probe-tip.

It is important not to let the probe tip slip between the pins and create a short-circuit.

LOGIC PULSER

If you have a board or a single chip and want to create activity (clock pulses), you can use a Logic Pulser. This piece of test equipment will produce a stream of pulses that can be injected into the clock-line (clock input) of a chip.

You can then use a Logic Probe at the same time on the outputs to observe the operation of the chip.

You can also use the Mini Bench Amplifier to detect "noise" or activity on the inputs and outputs of digital chips.

This only applies if the frequency is in the audio range such as scanning a keyboard or switches or a display.

This is how to approach servicing/testing in a general way. There are thousands of digital chips and if you want to test a specific chip for its exact performance, you will need to set-up a "test-bed."
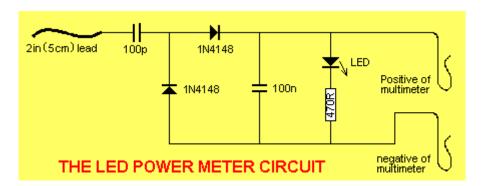
REMOTE CONTROLS

There are two types of remote control - Infrared and RF. Infrared is used for short-range, line-of-sight for TV's DVD's etc.
A few faults can be fixed, but anything complex needs a new remote control.
Check the batteries and battery-contacts. See if the IR LED is illuminating by focusing it into a digital camera and looking on the screen for illumination.
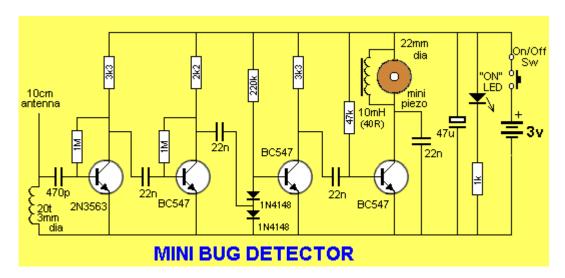The only other things are a sticky button, a worn-out button or a crack in the PC board. Water damage is generally too much work to repair.
RF remote controls for cars, garage doors etc need a second working unit to check the power output.
Here is a simple circuit that can be connected to an analog multimeter to detect the signal strength at a very close range:

THE LED POWER METER CIRCUIT

To hear the tone from a transmitter, the **Mini Bug Detector** circuit can be used:

MINI BUG DETECTOR

Any further investigation requires a circuit diagram so you can work out what is actually being sent from the transmitter.
Most of the time it is a faulty switch, battery or contacts. Make sure the setting is correct on the "dip switches" and use a working unit to compare all your testing.

TESTING VOLTAGES ON (in) A CIRCUIT

There are basically two different types of circuit.
1. **ANALOGUE CIRCUIT**
An analogue circuit can also be called an AUDIO CIRCUIT and the voltages at different points in a circuit can be measured with a multimeter but the changes (the waveforms) will be quite small or changing at a rapid rate and cannot be detected by a multimeter.
You need a CRO to "see" the signals or a **Signal Injector** to inject a waveform into the circuit and hear the result on the circuit's speaker.

2. **DIGITAL CIRCUIT**
A digital circuit can also be called a "Computer Circuit" or "Logic Circuit" and some of

the voltages can be measured with a multimeter (such as supply voltages) but the "signal lines" will be be changing from HIGH to LOW to HIGH very quickly and these signals are detected with a **Logic Probe**.

Here are some circuits with details of how to test the voltages.
Most circuits do not show voltages at various different points and we will explain what to expect on each "stage."

A "STAGE"

A stage is a set of components with an input and output. A "stage" can also be called a "Building Block."
Sometimes it has a capacitor on the input and one on the output.
This means the stage is completely isolated as far as DC is concerned.
The stage has a supply (a DC supply) and it is producing its own voltages on various points on the "stage." It can only process (amplify) "AC." (signals).
Sometimes the stage can be given a name, such as small-signal amplifier, push-pull amplifier or output.
If the stage has a link or resistor connected to a previous stage, the previous stage will have a "DC effect" on the stage. In other words it will be biasing or controlling the voltages on the stage. The stage may be called a "timer" or "delay" or "DC amplifier."

It is important to break every circuit into sections. This makes testing easy. If you have a capacitor at the input and output, you know all the problems lie within the two capacitors.
In a digital circuit (no capacitors) you need to work on each IC (integrated Circuit) and test the input for activity and all the outputs.

Once you have determined if the circuit is Analogue or Digital, or a combination of both, you have to look at the rail voltage and work out the size or amplitude of the voltage or waveform.
This is done before making a test, so your predictions are confirmed.
You will need a **multimeter** (either Digital or Analogue) a **Logic Probe** and a **Signal Injector** (**Tone Generator**). An analogue meter has the advantage that it will detect slight fluctuations of voltage at a test-point and its readings are faster than a digital meter. A digital meter will produce an accurate voltage-reading - so you should have both available.

HIGH IMPEDANCE AND LOW IMPEDANCE

Every point in a circuit has a characteristic called "IMPEDANCE." This has never been discussed before in any text book. That's why it will be new to you.
In other words, every point will be "sensitive to outside noise."
An audio amplifier is a good example. If you put your finger on the active input, it will produce hum or buzz in the speaker. This is because it is a HIGH IMPEDANCE line or high impedance section of the circuit.
The same applies to every part in a circuit and when you place Test Equipment on a line for testing purposes, the equipment will "upset" the line. It may be very slight but it can also alter the voltage on the point CONSIDERABLY.
We have already mentioned (above) how a cheap multimeter can produce a false reading when measuring across a 1M resistor. That's why you need high impedance test Equipment so you do not "load" the point you are testing and create an inaccurate reading.
The word **Impedance** really means resistance, but when you have surrounding components such as diodes, capacitors, transistors, coils, Integrated Circuits, supply-voltages and resistors, the combined effect is very difficult to work out as a "resistance" and that's why we call it "Impedance."
The term "**High and Low Impedance**" is a relative term and does not have any absolute values but we can mention a few points to help you decide.
In general, the base of a transistor, FET input of an IC are classified as HIGH IMPEDANCE.
The output of these devices are LOW IMPEDANCE.

Power rails are LOW IMPEDANCE.
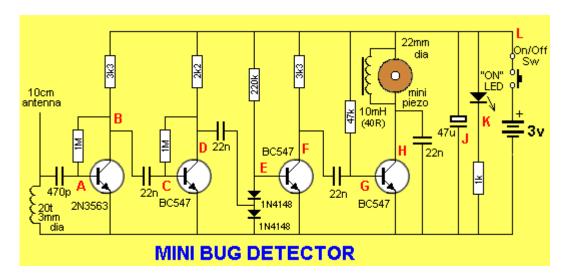An oscillator circuit and timing circuit are HIGH IMPEDANCE.
A LOAD is low impedance.
And it gets tricky: An input can be designed to accept a low-impedance device (called a transducer or pick-up) and when the device is connected, the circuit becomes LOW impedance, but the input circuitry is actually high impedance.
The impedance of a diode or LED is HIGH before the device sees a voltage higher than the junction voltage and then it becomes LOW Impedance.
Impedance is one of the most complex topics however it all comes down to testing a circuit without loading it.
That's why test equipment should have an input impedance higher than 1M.

The first circuit we will investigate is the **Mini Bug Detector**, shown above and below. Points on the circuit have been labelled A, B, C etc:

MINI BUG DETECTOR

Point A - The first transistor is "self-biased" and will have 0.6v on the base. The antenna is connected to a 20 turn coil and you might think the coil will "short" the signals to earth.
But the coil and 470p capacitor form a circuit that oscillates at a high frequency when the antenna wire picks up stray signals. The coil and capacitor actually amplify the signals (see Talking Electronics website: Spy Circuits to see how a TANK CIRCUIT works) and these signals enter the base of the first transistor.
This is classified as a HIGH Impedance section because the signals are small and delicate and any loading via test equipment will kill them. The first transistor amplifies the signals about 70 times and they appear at **Point B**.

The signal passes though a 22n to **Point C** and the transistor amplifies the signal about 70 times to **point D**. **Point C** is classified as high impedance as any voltage measurement at this point will upset the biasing of the stage as a few millivolts change in base-voltage will alter the voltage on the collector considerably. **Point D** is classified as low impedance as any voltage-testing will not alter the voltage appreciably.
The output of the second stage passes through a capacitor to the join of two diodes. These two diodes are not turned on because the voltage at **Point E** can never rise above 0.7v as this is the voltage produced by the base-emitter of the third transistor.
The purpose of the two diodes is to remove background noise. Background noise is low amplitude waveforms and even though the transistor is turned on via the 220k, low amplitude signals will not be received. The third transistor works like this: It cannot be turned ON any more because any waveform from the 22n will be "clipped" by the bottom diode and it will never rise above 0.6v.
So, the only signal to affect the transistor is a negative signal - to turn it OFF.
Firstly we have to understand the voltage on the 22n. When the second transistor is sitting at mid-rail voltage, the 22n gets charged via the 2k2 and lower diode. When

the transistor gets tuned ON, the collector voltage falls and the left side of the 22n drops. The right side of the 22n also drops and when it drops 0.6v, the top diode starts to conduct and when the voltage on the 22n drops more than 0.6v the third transistor starts to turn OFF. This effect is amplified by the transistor at least 100 times and appears at **Point F.** All the voltages around the two diodes are classified as HIGH Impedance as any piece of test equipment will upset the voltage and change the output.

There are some losses in amplitude of the signal as it passes through the 22n coupling capacitors but the end result is a very high strength signal at **point G.** The 4th transistor drives a 10mH choke and the mini piezo is effectively a 20n capacitor that detects the "ringing" of the inductor to produce a very loud output.

The 22n capacitor on the collector eliminates some of the background noise. The choke and piezo form an oscillatory circuit that can produce voltages above 15v, even though the supply is 3v.

The 47n capacitor at **Point J** is to keep the supply rails "tight" (to create a LOW Impedance) to allow weak cells to operate the circuit.
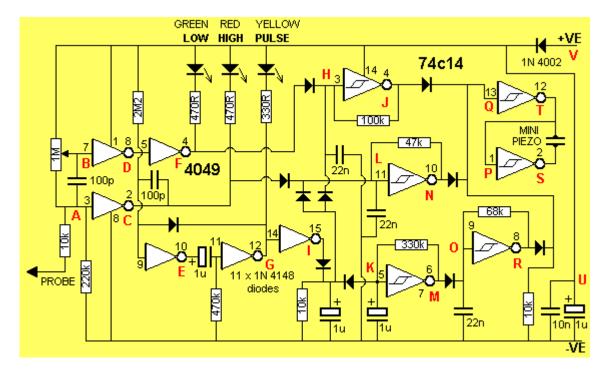
The "Power-ON" LED tells you to turn the device off when not being used and **Point L** is the power supply - a low impedance line due to the 47u electrolytic.

Testing the Mini Bug Detector
To test the Mini Bug Detector, you will need a Signal Injector.
Place the Injector on **Point G** and you will hear a tone. Then go to **E**, **C** and **A**. The tone will increase in volume. If it does not increase, you have pin-pointed the faulty stage.

The next circuit is a combination of digital and analogue signals. It is a **Logic Probe**:

The voltage on a circuit (to be tested) is detected by the probe at **Point A** of the circuit above and the "tip" is classified as "reasonably high impedance" as it has a 220k resistor between the tip and 0v rail. The 1M reduces the impedance by about 20% but the inputs of the two inverters have no effect on the "tip" impedance as they are extremely high input-impedance devices.

The 1M trim pot is designed to put put a voltage on **point B** that is slightly higher than mid-rail so the green LED is turned off.

Point A will see a voltage below mid-rail and **point C** will be HIGH. **Point C** and **F** are low-impedance outputs.

When the tip of the probe is connected to a LOW voltage, Point B sees a LOW and Point F goes LOW to illuminate the green LED. At the same time it removes the "jamming voltage" produced by the diode between pin 4 of the 4049 and pin 3 of the

74C14 and the oscillator between **points H and J** produces a low-tone via the 100k resistor and 22n to indicate a LOW.

When the probe tip sees a HIGH, a lot more things happen.

Point C goes LOW and turns on the red LED. At the same time the 100p is in an uncharged state and the right lead goes LOW. This takes the left lead LOW as the left lead connects to a HIGH Impedance line and pin 9 goes LOW. This makes **point E** HIGH

and since the 1u is in an uncharged state, pin 11 goes HIGH. This makes **point G** LOW and the diode between pins 9 and 12 keeps pin 9 LOW and takes over from the pulse from the 100p. The yellow LED is illuminated. The 1u starts to charge via the 470k and when it is approx half-charged, pin 11 sees a HIGH and **point G** goes low. This creates the length of pulse for the yellow LED.

At the same time, **Point L** goes LOW because the "jamming diode" from pin 2 of the 4049 goes low and allows the inverter between point L and N to produce a tone for the piezo.

In addition, **Point I** goes HIGH and quickly charges a 1u electrolytic. This removes the effect of the jamming diode on pin 5 of the 74C14 and a low frequency oscillator made up of 68k and 1u between pins 5&6 turns on and off an oscillator between **points O** and **R** to get a beep. The mini piezo is driven n bridge mode via the two gates between **points QT** and **PS**.

Point U is a 1u electrolytic to reduce the impedance of the power rail and **Point V** is a protection diode to prevent damage if the probe is connected to the supply around the wrong way.

Testing the Logic Probe

You can test the Logic Probe with the simple Logic Probe with Pulse project described above. It will let you know if each point in the circuit is HIGH or LOW. You will also find out the difficulty in testing the points that are HIGH Impedance, as the Probe will upset the voltage levels and the reading may be inaccurate.

More circuits will be added here in the future.

THE VOLTAGE DIVIDER - this topic could fill a book.

You need to read lots of other sections in this eBook, including the section on measuring across a resistor with a multimeter, and high impedance circuits, to fully understand the complexities of a VOLTAGE DIVIDER CIRCUIT.

It is one of the most important BUILDING BLOCKS to understand. Even though it may consists of two components, you have to understand what is happening between these two components. You have to realise there is a voltage at their join that will be rising and falling due to one of the components changing RESISTANCE. Sometimes you can work out the voltage at the join by using Ohm's LAW but quite often it will be impossible as it is changing (rising and falling) during the operation of the circuit.

At the beginning of this discussion we will only dealing with DC circuits and the voltage across a particular component will be due to its RESISTANCE. We are not going into any formulas, as it is very easy to measure the voltages with a multimeter set to VOLTS and you will have an accurate result.

The simplest two components in series are resistors. They always have the same resistance during the operation of a circuit and the voltage across each will not change.

In a further discussion we will cover "resistors" that change value according to the temperature. These are called THERMISTORS. And we have "resistors" that change value according to the light they receive. These are called LIGHT DEPENDENT RESISTORS (LDR's) or PHOTO RESISTORS.

A transistor that is partly or fully turned ON can be considered to be similar to a resistor.

In these 3 cases we need to measure the voltage at the join with a voltmeter as it will be a lot of work to measure the resistance and work out a value.

You can also keep a voltmeter on the joint and watch the voltage change.

Finally we have some components that produce a fixed voltage across them (or nearly fixed) and the remaining voltage is dropped across a resistor. These

components MUST have a resistor connected in series to limit the current and allow the component to pass the specified in the datasheet.

These devices include LEDs, diodes and zener diodes. A LED will have a fairly fixed voltage across it from 1.7v to 3.6v depending on the colour. A diode will have a voltage of 0.7v across it when it is connected to a voltage via a resistor. And a zener diode will have a fixed voltage across it when it is connected with the cathode to the positive rail via a resistor. The voltage across it will be as marked on the zener.

The concept of a VOLTAGE DIVIDER is very simple, but it takes a lot of understanding because both VOLTAGE and CURRENT are involved in the UNDERSTANDING-PROCESS.

Each component has a resistance and this can be measured with a multimeter. When two components are connected in series, a current will flow and a voltage will develop across each item.

More voltage will develop across the item with the higher resistance and the addition of each voltage will always equal the supply voltage.

That's the simple answer.

There is a little more involved . . . It is the word CURRENT. Here is an explanation: Suppose we have a 1k and 2k resistor on a 12v supply. The voltage at the join will be 4v.

In other words, there will be 4v across the 1k and 8v across the 2k.

If we have a 10k and 20k resistors in series, the voltage will also be 4v at the join.

If we have a 100k and 200k resistors, the voltage will also be 4v at the join.

The voltage will be the same in all cases, but the current will be different. The current in the second case will be one-tenth and only one hundredth in the third case.

If you want to go further, place a one ohm and two ohm in series and get 4v. But the resistors will get very hot and burn out very quickly.

SOLDERING

Here are three 30-minute videos on soldering.
1. **TOOLS**
2. **Soldering components**
3. Soldering **SURFACE MOUNT** components

TESTING A MOTOR

Strictly speaking, a motor is not an electronic component, but since a website gave a useless description on testing motors, I have decided to supply the correct information.

The only REAL way to test a motor is to have two identical motors and check the torque by connecting them to a low voltage and trying to stop the shaft with your fingers. This will give you two results. Firstly it will let you know the torque of the motor.

This is the twisting effect of the shaft. There is no way to determine the torque by knowing the voltage or current.

The unknown factor is the strength of the field magnets (permanent magnets) and this determines the torque.

Secondly, feeling the shaft will let you know if the torque is even for a complete revolution.

By having two identical motors, you can see if one has a lower torque.

Almost nothing can go wrong with a motor except for the brushes. If the brushes wear out, additional resistance will be produced at the interface between the brush and commutator and this can be detected by allowing the shaft to rotate slowly and feeling the resistance as it revolves. A 3-pole motor will have three places where the strength is greatest and each should have the same feeling. A 5-pole motor will have five places of strength.

If the strength is weak or not uniform, the motor is faulty.

You cannot test a motor with a multimeter as the resistance of the armature winding is very low and if the motor is allowed to spin, the back voltage produced by the spinning, increases the reading on the meter and is false.

Micro motors have a coreless armature. This means the 3 windings for the armature are wound on a machine then bent slightly into shape and glued. A circular magnet with 3 poles is in the centre and the armature rotates around this.
This type of motor is reasonably efficient because the armature is the greatest distance from the point of rotation, and the motor reaches full RPM very quickly because the armature has very little inertia.
I have not heard of the armature-winding flying apart but if you hear any scraping noise, it may be the winding.
3-pole, 5-pole and micro motors can be found in printers, eject mechanisms of CD players, toys, RC helicopters, cars etc and rarely fail.
Motors do not work on "voltage." They actually work on CURRENT and as you increase the voltage, more current will flow and produce a stronger magnetic field (by the winding on each pole). This magnetic field will be attracted by the permanent magnet surrounding the armature and repelled by the surrounding permanent magnet, depending on where the face of the pole is, during each revolution.
If the permanent magnet is not very strong, the repulsion part of the interaction will be very weak and thus the torque will be small.
Because motors work on "current" you must have a high current available when you increase the voltage as the motor will require short bursts of high current during each revolution.
It is the combination of voltage and current (called watts) that gives the motor "strength" (torque) as well as the "strength" of the permanent magnets (called the field magnets) and the number of turns of wire on each pole (and the gauge of wire). Basically, if a motor is hard to spin, and has 3 "hard spots" on each revolution, it will be powerful.
A 2-pole motor does not self-start and will spin in either direction. But a 3-pole motor will self-start and you can determine the direction of rotation.
A 5-pole motor has a lower RPM. It is slightly smoother in output but is not more powerful than a 3-pole version.
A motor with "permanent magnets" is called a DC motor as it will not work on AC. If the magnets are replaced with a coil, it will work on AC and it will be called a "shunt wound" motor of the field coil is connected across the same terminals as the brushes or a "series wound" motor if the field coil is in series with the armature.

TESTING COMPONENTS "IN-CIRCUIT"
You can test components while they are IN CIRCUIT, but the surrounding components will have an effect on the results.
You can get all sorts of "In-Circuit" testers. They are expensive and offer little more accuracy than a multimeter.
In-Circuit testing with a multimeter can give you the same results as a tester.
All you have to do is turn the project ON and use a multimeter (set to voltage) to determine the voltage at various points. It is best to have a circuit of the equipment so you can what to expect at each point.
Only major departures from the expected can be located in this way.
Obviously the first thing to look for is burnt-out components. Then feel components such as transistors for overheating.
The look for electrolytics that may be dry. Sometimes these have changed colour or are slightly swollen.
If they are near hot components, they will be dry.
For the cost of a few dollars I change ALL THE ELECTROLYTICS in some pieces of equipment, as a dry electrolytic is very difficult to detect.
Testing a transistor "in-circuit" is firstly done with the supply ON. That's because it is quicker.
Measure the voltage between ground and collector.
In most cases you should get a voltage of about half-rail. If it is zero, or close to rail voltage, you may have a problem.
Turn off the supply and use the multimeter on low-ohms to measure all six resistances between the leads.
A low resistance in both directions on two leads will indicate a fault.
Resistors almost NEVER go "HIGH." For instance, a 22k will never go to 50k.
However a low-value resistor will "burn-out" and you will read the value of the surrounding components.

Don't forget, some low-value resistors are designed to burn-out (called fusible resistors) and anytime you find a damaged low-value resistor, you will need to look for the associated semiconductor.

You can replace the resistor quickly and turn the circuit ON to see it burn out again. Alternatively you can trace though the circuit and find the shorted semiconductor. It's always nice to "see the fault" then "fix the fault."

Sometimes a transistor will only break-down when a voltage is present, or it may be influenced by other components.

When the piece of equipment is turned OFF, you can test for resistance values. The main thing you are looking for is "dry joints" and continuity. Dry joints occur around the termination of transformers and any components that get hot. Rather than wasting time checking for dry joints, it is better to simply go over the connections with a hot iron and fresh solder.

You may need to check the continuity of a track (trace) and it may go from one side of the PC board to the other.

Use a multimeter set to low-ohms and make sure the needle reads "zero-ohms."

It is very dangerous to do any testing on a project using a multimeter set to "amps" or "milliamps."

You cannot test "current flowing through a component" by placing the probes across a component. You will simply over-load the rest of the circuit and create a problem.

To find out if current is flowing though a circuit or a low-value resistor, turn the project ON and measure the voltage either across the component or the voltage on one end then the other.

A voltage-drop indicates current is flowing.

That's about it for testing "in-circuit." Use the rest of this eBook to help you with diagnosis.

Don't think an IN-CIRCUIT COMPONENT TESTER is going to find a fault any faster than a multimeter. They all use a multimeter principle.

SHORT CIRCUIT

Nearly every component can fail and produce an effect called a SHORT CIRCUIT. This basically means the component takes more current than normal and it may fail completely or simply take more current and the operation of the circuit may be reduced only a small amount.

The resistance of the component may reduce a very small amount but this may have a very large effect on the operation of the circuit.

For instance, two turns in the horizontal or vertical winding of a yoke on the picture tube or monitor may arc and weld together and reduce the size of the picture on the screen, but measuring the winding will not detect the difference in resistance.

The same with the windings on a motor and a short between two winding in a transformer.

If the "short" is between two near-by turns, the change in resistance will be very small. If the "short" is between to different layers, the resistance will be reduced and it may be detected.

When a "short" occurs, the winding turns into a transformer. To be exact, an AUTO-TRANSFORMER.

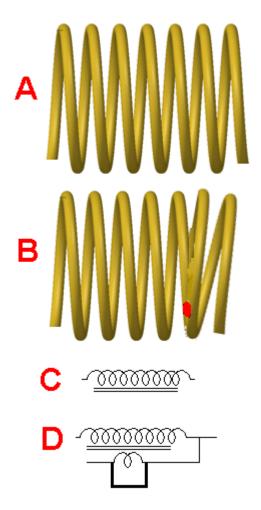In the following diagram you can see a normal winding in fig A:

Fig B shows two turns touching each other and if the wire is enamelled, the coating has been damaged so the copper wire from the two turns is touching. This is called a SHORTED TURN.

In fig C you can see two turns touching.

In fig D the shorted-turn has been moved to the other side of the symbol to show the effect it has on the operation of the winding.

The shorted-turn is exactly like the secondary of a transformer with a "jumper" across the output.

This will produce a very high current in the secondary.

A very high current flows through the shorted turn and this changes the operation of the rest of the winding.

1. In most cases a SHORT CIRCUIT can be detected by feeling the additional heat generated by the component.

2. Next, turn off the supply and measure the resistance of the component. If it is lower than expected, the component will be faulty.

3. Next, measure the voltage across the component. If it is lower than normal, the component will be faulty.

4. Next, measure the current taken by the component. If it is higher than normal, the component will be faulty.

5. If the component is an inductor, such as a motor, coil or transformer, you can use an inductance meter. Compare a good winding with a faulty winding. Sometimes the fault will disappear because an arc develops across the fault when the component is operating.

INTERNAL AND EXTERNAL SHORTS

An **internal short** refers to two windings shorting together and the winding has a very high resistance between the winding and the frame on which it is wound. An external short refers to a winding shorting to the frame of the component - such as one of the armature windings shorting to the metal core, around which the wire is

wound.

This may not be important unless another winding shorts to the metal frame and creates "**inter winding**" problems (**inner winding** problems is within the same winding).

The opposite to a short circuit is an OPEN CIRCUIT.

This is generally a broken lead or contact or a wire that has "burnt-out" or been "eaten-away" by acid attack or galvanic action by water and voltage (current).
1. No current will flow when an OPEN CIRCUIT exists.
2. The voltage on each end of the OPEN CIRCUIT will not be the same.
3. Measure the current across the OPEN CIRCUIT and determine if excess current is flowing.
4. Join the two ends of the OPEN CIRCUIT and see if the circuit operates normally.

> **HEATSINKS**

This is not an electronic component but it can certainly affect the operation of a circuit.

If you cannot hold your fingers on a heatsink, it is getting too hot. This is because the actual location where the heat is being generated is much hotter than the part you are touching.

Transistors and IC's can withstand a high temperature but if they go above this temp, they BLOW UP.

They also have a shorter life when operating at a high temperature.

The secret to a good heatsink is called an INFINITE HEATSINK.

This is the metal frame of a case.

There are lots of charts and data on choosing a heatsink but they don't take into account two factors:

Sometimes a circuit takes a very high current for a short time and this creates a high temperature gradient. This will cause the transistor to get very hot and fail.

The solution is to have two or more transistors in parallel to separate the "heat spots."

The second problem with designing a heatsink is the unknown location of the heatsink and the air-flow. Products placed on a shelf or in a cupboard will get very little air-flow.

Remember: some transistors are mounted on thermal insulators. This means the transistor will have a voltage on it but the heatsink will be zero voltage.

The temperature of the transistor will be MUCH HIGHER than the heatsink under the transistor and the transfer of the heat from the transistor to the heatsink will be very slow. This can be the cause of the transistor failing. Sometimes the transistor will fail because insulation is high temp plastic and it gets brittle. The plastic can carbonise and leak and sometimes a voltage can flash through the insulator. Some amazing things have happened under these transistors and you may need to pull it apart and replace all the insulation.

Finally, feel the heatsink after 15 minutes and feel right up to the transistor. If you cannot touch the transistor, increase the thickness of the heatsink or use two transistors to dissipate the heat.

To design a heatsink, you have to have some idea of the size of a heatsink for the application.

Charts and data can send you in the wrong direction.

Start with a heatsink twice the recommended size and feel the temp after 15 minutes. Put the project in a cupboard and see how the temperature rises.

If possible, connect the heatsink to the metal case to get added dissipation and if you include fan-cooling, remember the fan will eventually gather dust and reduce its efficiency.
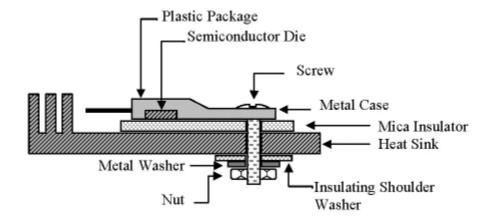
It is very difficult to explain how heat passes through a mica washer or plastic washer, but if the transistor has a copper base, the heat transfer has a value of 400. For aluminium it is 200. If it is steel, the transfer has a value of 50. For a mica sheet it is 1 and for plastic it is 0.1

Even though the sheet is very thin, the transfer is a lot less than metal-to-metal transfer.

Most references state the temperature difference is about one degree C for each watt of heat generated by the transistor.
Don't believe anything you read.
Feel the temperature yourself and if you cannot hold your finger on the transistor, fix the problem.

In the end, use a heatsink 50% larger than recommended.
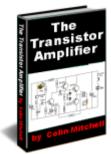
THE END

This is not the full story to learning about servicing. It is just the beginning.
We have only covered the simplest tests and shown how 90% of faults can be found by checking voltages, waveforms and looking for obvious things such as burnt out components, cracks in PC boards.
The author has fixed over 35,000 TV's, radios, stereos, VCRs and all those things that were on the market 30 years ago.
Things have not changed. It's just that some repairs cost nearly as much as buying a new product and half the customers opt for dumping a faulty item and buying the latest "flat screen" version. That's why you have to get things through the workshop as fast and as cheaply as possible, to make a living.

If you want any more devices added to this list, email **Colin Mitchell**.

To help with understanding how a transistor circuit works, we have produced an eBook: **The Transistor Amplifier**. It covers a whole range of circuits using a transistor.

Not copyright by Colin Mitchell
You can use any of this material. Please pass this eBook to your friends and let them know that everything on the web is FREE. I have looked at all the "Pay Sites" and found the information they "sell" is available on the web at NO COST.
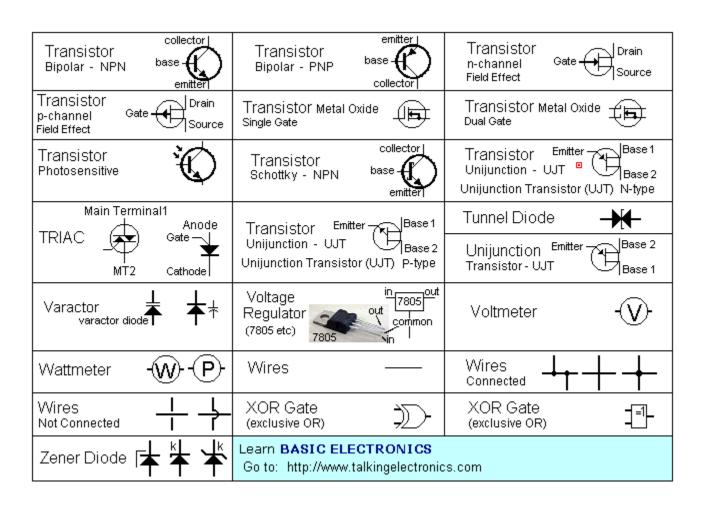Nearly all text books are also downloadable for free on "Download.com" etc and when you see a used copy of a $74.00 textbook on Amazon for $12.00 you realise many users have already discarded their copy. A good textbook never gets thrown out or sold for $12.00!!!

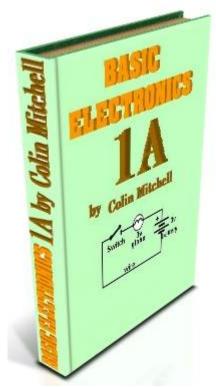**See the enormous amount of information on Talking Electronics website**

CIRCUIT SYMBOLS by TALKING ELECTRONICS

AC current: voltage:	ALTERNISTOR TRIAC — A TRIAC and 33 - 43v DIAC — Main Terminal1, Gate, Main Terminal 2	Ammeter (amp meter)
AND Gate	AND Gate	Antenna balanced
Antenna Loop, Shielded	Antenna Loop, Unshielded	Antenna unbalanced
Attenuator, fixed (see Resistor)	Attenuator, variable (see Resistor)	Battery
Bilateral Switch (DIAC)	Bridge Rectifier (Diode Bridge)	BUFFER (Amplifier Gate)
BUFFER (Amplifier Gate)	Buzzer	Capacitor feedthrough
Capacitor non-polarised	Capacitor polarised (see electrolytic)	Capacitor Variable
Cavity Resonator	Cell	Circuit Breaker
Coaxial Cable	CRO – Cathode Ray Oscilloscope	Crystal Microphone (Piezoelectric)
Connectors — Plug (male), Jack (female), connected — Plug (male), (female)	Crystal Piezoelectric	DC current: voltage:
	Darlington Transistor — collector, base, emitter	Delay Line
	DIAC (Bilateral Switch)	Diode k
Diode - Gunn	Diode - Light Emitting (LED)	Diode Photo Sensitive
Diode Photovoltaic λ	Diode Bridge (Bridge Rectifier)	Diode - Pin
Diode - Varactor	Diode - Zener	Earth Ground
Earpiece (earphone, crystal earpiece)	Electroluminescence el	Electret Microphone (Condenser mic)
Electrolytic (Polarised Capacitor) — alternate symbols: (positive on top)	Electrolytic - Tanatalum — positive end, black band or chamfer, 10u tantalum	Exclusive-OR Gate (XOR Gate)
		Exclusive-OR Gate (XOR Gate)
Field Effect Transistor (FET) n-channel also: N-Channel J FET — Gate, Drain, Source	Field Effect Transistor (FET) p-channel also: P-Channel J FET — Gate, Drain, Source	Flashing LED (Light Emitting Diode) (Indicates chip inside LED)

Ferrite Bead	Fuse	Galvanometer
Globe	Ground Chassis	Ground Earth
Heater (immersion heater) (cooker etc)	IC Integrated Circuit — power — ground	Inductor Air Core
Headphone		Inductor Iron Core or ferrite core
Inductor Tapped	Inductor Variable	Integrated Circuit
Inverter (NOT Gate)	INVERTER (NOT Gate)	
Jack Co-axial	Jack Phone (Phone Jack)	Jack Phone (Switched)
Jack Phone (3 conductor)	Key Telegraph (Morse Key)	Lamp Incandescent
Lamp - Neon	LASCR (Light Activated Silicon Controlled Rectifier)	LDR (Light Dependent Resistor)
LASER diode laser diode photo diode	Light Emitting Diode (LED)	Light Emitting Diode (LED - flashing) (Indicates chip inside LED)
Mercury Switch	Micro-amp meter (micro-ammeter)	Microphone (see Electret Mic)
Microphone (Crystal - piezoelectric)	Milliamp meter (milli-ammeter)	Motor
NAND Gate	NAND Gate	Nitinol wire "Muscle wire"
Negative Voltage Connection	NOR Gate	NOR Gate
NOT Gate Inverter	NOT Gate Inverter	Ohm meter
Operational Amplifier (Op Amp)	Optocoupler (Transistor output)	Opto Coupler (Opto-isolator) Photo-transistor output
Optocoupler (Darlington output)	Opto Coupler (Opto-isolator) TRIAC output	OR Gate
OR Gate	Oscilloscope see CRO	Outlet (Power Outlet)
Piezo Diaphragm	Photo Cell (photo sensitive resistor)	Photo Diode
Photo Darlington Transistor	Photo FET (Field Effect Transistor) Gate Drain Source	Photo Transistor

Photovoltaic Cell (Solar Cell)	**Piezo Tweeter** (Piezo Speaker)	**Positive Voltage** Connection
Potentiometer (variable resistor)	**Programmable Unijunction Transistor** PUT — gate, anode, cathode	**Rectifier** Silicon Controlled (SCR) — Anode, Gate, Cathode
Rectifier Semiconductor — k	**Reed Switch**	**Relay** - spst
Relay - spdt	**Relay** - dpst	**Relay** - dpdt
Resistor Fixed	**Resistor** Non Inductive	**Resistor** preset
Resistor variable	**Resonator** 3-pin	**RFC** Radio Frequency Choke
Rheostat (Variable Resistor)	**Saturable Reactor**	**Schmitt Trigger** (Inverter Gate)
Schottky Diode (also Shottky) — k, k. Low forward voltage 0.3v. Fast switching. also called Schottky Barrier Diode	**Shielding** - - - - - **Signal Generator**	**Shockley Diode** — k. 4-layer PNPN device. Remains off until forward current reaches the forward break-over voltage.
Silicon Bilateral Switch (SBS) T₂ Terminal, Gate, T₁ Terminal — e.g: BS08D — T₂ G T₁	**Silicon Unilateral Switch (SUS)** Anode, Gate, Cathode(k) — A G k	**Silicon Controlled Rectifier (SCR)** Anode, Gate, Cathode **Solar Cell**
Surface Mount SOT-23 — c b e, & LED, no connection, A k	**Switch** - spst **Switch** - spdt **Switch** - dpst **Switch** - dpdt **Switch** - mercury tilt switch **Spark Gap**	**Switch** - process activated — normally open: normally closed: — Flow, Level, Pressure, Temperature **Speaker** 8R
Switch - push (Push Button)	**Switch** - push off (used in alarms etc)	**Switch** - Rotary
Test Point — o **Thermal Probe** NTC t°= ? **NTC:** as temp rises, resistance decreases	**Thyristors:** Main Terminal1, Bilateral Switch, Anode, Gate, Cathode, Anode, MT2, Gate, Cathode — DIAC SCR TRIAC TRIAC	**Thermocouple** **Tilt switch** mercury **Touch Sensor**
Transformer Air Core	**Transformer** Iron Core	**Transformer** (Tapped Primary/Sec)

Transistor Bipolar - NPN collector, base, emitter	**Transistor** Bipolar - PNP emitter, base, collector	**Transistor** n-channel Field Effect Gate, Drain, Source
Transistor p-channel Field Effect Gate, Drain, Source	**Transistor** Metal Oxide Single Gate	**Transistor** Metal Oxide Dual Gate
Transistor Photosensitive	**Transistor** Schottky - NPN collector, base, emitter	**Transistor** Unijunction - UJT Emitter, Base 1, Base 2 Unijunction Transistor (UJT) N-type
TRIAC Main Terminal1, MT2 Gate, Anode, Cathode	**Transistor** Unijunction - UJT Emitter, Base 1, Base 2 Unijunction Transistor (UJT) P-type	**Tunnel Diode**
		Unijunction Transistor - UJT Emitter, Base 2, Base 1
Varactor varactor diode	**Voltage Regulator** (7805 etc) in, out, common, 7805	**Voltmeter**
Wattmeter	**Wires**	**Wires** Connected
Wires Not Connected	**XOR Gate** (exclusive OR)	**XOR Gate** (exclusive OR)
Zener Diode k	Learn **BASIC ELECTRONICS** Go to: http://www.talkingelectronics.com	

More chapters of this eBook on: Talking Electronics.com

For any enquiries email **Colin Mitchell**

BASIC ELECTRONICS
(this is the Basic Electronics section i.e. Page 1)
(Chapters 1 and 3 are available as .pdf)

Quick Quiz - to see how much you know
Encyclopedia of Components - this is excellent !!!

Remember: **the animations do not work in .pdf**
the site is being constantly updated

INDEX

KIT OF PARTS

Talking Electronics supplies a kit of parts that can be used to build the majority of the circuits in this eBook.
The kit costs $15.00 plus postage.

Kit for Transistor Circuits - $15.00

BUY NOW

A kit of components to make many of the circuits described in this eBook is available for $15.00 plus $7.00 post.
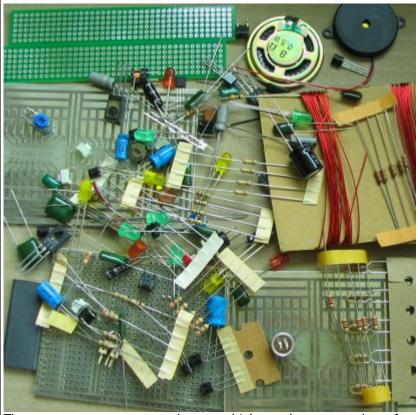Or email Colin Mitchell: talking@tpg.com.au

The kit contains the following components:
(plus **extra** 30 resistors and 10 capacitors for experimenting), plus:

3 - 47R
5 - 220R
5 - 470R
5 - 1k
5 - 4k7
5 - 10k
2 - 33k
4- 100k
4 - 1M
1 - 10k mini pot
1 - 100k mini pot
2 - 10n
2 - 100n
5 - 10u electrolytics
5- 100u electrolytics

5 - 1N4148 signal diodes
6 - BC547 transistors - NPN - 100mA
2 - BC557 transistors - PNP - 100mA
1 - BC338 transistor - NPN - 800mA
3 - BD679 Darlington transistors - NPN - 4amp
5 - red LEDs
5 - green LEDs
5 - orange LEDs
2 - super-bright WHITE LEDs - 20,000mcd
1 - 3mm flashing LED
1 - mini 8R speaker
1 - mini piezo
1 - LDR (Light Dependent Resistor)
1 - electret microphone
1m - 0.25mm wire
1m - 0.5mm wire
1 - 10mH inductor
1 - push button
5 - tactile push buttons
1 - Experimenter Board (will take 8, 14 and 16 pin chips)
5 - mini Matrix Boards: 7 x 11 hole,
 11 x 15 hole, 6 x 40 hole, surface-mount 6 x 40 hole board and
others.

Photo of kit of components.
Each batch is slightly different:

There are more components than you think. . . plus an extra bag of
approx 30 components. The 8 little components are switches and the
LDR and flashing LED is hiding.
In many cases, a resistor or capacitor not in the kit, can be created
by putting two resistors or capacitors in series or parallel or the next
higher or lower value can be used.

BEFORE WE START

Too many text books start with the physics of the atom and have equations and mathematics to show how smart the author is.

Don't worry, we wont have any physics or equations.

The reason . . .

This is not a physics course. It is a practical electronics course to teach the basics as quickly as possible. There are no equations because most transistor circuits cannot be worked out mathematically as the gain of a transistor changes according to the current-flow and these gain-values are never provided. So the mathematics is worthless.

To get an answer, all you have to its build the circuit and measure the values with a multimeter.

Also lots of discussions in text books will never be used in your next 40 years of electronics, so this course doesn't have any unnecessary material and is much-more concentrated than anything you have read before.

Every frame contains important points - especially the animations - as they show you how a circuit works in slow-motion - something that has NEVER been done before.

ELECTRONICS BLOCKS

Here is an idea from Instructables to produce blocks with screws, containing a single component and they can be connected with jumper leads (alligator clips).

SLOTTED HEAD

PHILIPS HEAD

Use a slotted head for the negative screw and a philips head for the positive screw.

Learn electronics from the beginning . . .

START HERE:
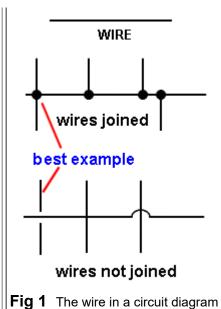
All electrical and electronic components need wire to connect them to the circuit.

In a diagram called a CIRCUIT DIAGRAM, the wires are drawn as lines.

When the wires (or lines) cross, they may be joined or just passing.

It is **VERY IMPORTANT** to show the difference between lines that are **JOINED** and lines that are **NOT JOINED**.

When the lines are joined, it is best to place a dot

WIRE

wires joined

best example

wires not joined

Fig 1 The wire in a circuit diagram

on the connection to PROVE the lines are joined. When the lines are just crossing, a gap should be made so it is obvious that one wire goes under the other and does not touch.
Lines should be "across the page" or "up and down." Very few lines should be at 45°.
You can make a line thicker to indicate a power rail or a wire that will be thick in reality.
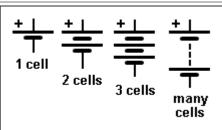
to Index

1 cell

2 cells

3 cells

many cells

Fig 2 A single Cell and many Cells

Next we need a **battery**. A battery consists of two or more cells. The positive terminal of a battery is the long line in the diagram and you must add the voltage (of the cell or battery) to the symbol as a single cells can be 1.2v, 1.5v, 2.2v or up to 3.6v.
The symbol does not let you know the voltage.
The positive is always at the top and is the longest line on the battery symbol.

to Index

globe

Fig 3: A Globe

Next we need a globe. A globe has two connections (a fine wire inside a glass bulb glows when the globe is connected to a battery).

to Index

globe

battery

wire

3v
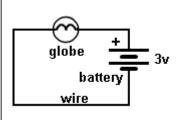
Fig 4: A Circuit

With a globe, battery and wire we have produced a **CIRCUIT**. A **CIRCUIT** is a complete path and we say the "electricity" the **CURRENT** emerges from the positive of the battery, moves through the globe and returns to the battery via the wire.
If the globe is a "3v GLOBE" it will glow when connected to a 3v battery.
The globe can be connected either way around.
The circuit we have shown is called a **SCHEMATIC** and consists of symbols: a globe symbol and a battery symbol.
The line connecting the two components is called WIRE.

to Index

Switch

3v globe

3v battery

wire

on   off

not used

off   on

SWITCH

Fig 5:Adding a SWITCH
To turn the globe ON and OFF we need a SWITCH.
The switch may be a push button, a toggle switch (a "click" action) or a slide switch. You could

twist the wires together and untwist them. The result is the same. We say the circuit is "broken" or "open" via the switch and the lamp does not glow. Closing the switch turns ON the globe.

Fig 6: Naming the RAILS

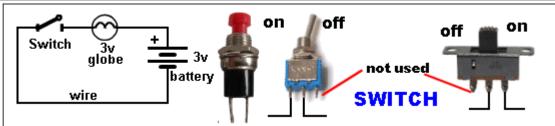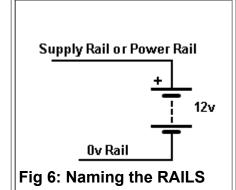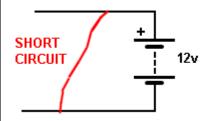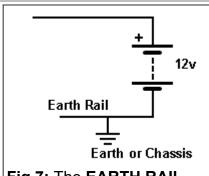The top rail of the **CIRCUIT DIAGRAM** is called the **SUPPLY RAIL** or **POWER RAIL**.
The lower rail is called the **0v Rail** or **EARTH RAIL**.
Do not connect the Supply rail (+12v) to the 0v rail as this will cause a high current to flow and is called a **SHORT CIRCUIT**:

Fig 7: The EARTH RAIL
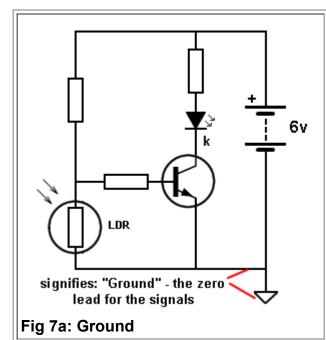
The lower rail is also called the **Chassis**.
This comes from the "old days" when electronics constructors build radios on a metal chassis (metal box) and it was connected via wire to a pipe in the ground to help the radio pick up distant radio stations.
The term also comes from car and truck wiring where one side of each globe is connected to the frame or chassis so that only one wire is needed to each globe and the return "path" is via the chassis.

Fig 7a: Ground

signifies: "Ground" - the zero lead for the signals
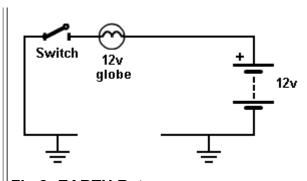
Some circuits identify the "**Ground Lead**" or "**Ground Wire**" of a project to show where all the **signals** have been "referenced to." In other words, all the signals rise and fall above and below this "Ground wire" or "Ground Lead." This lead may not be at **earth potential** as the project may be in a plastic box but it identifies where the earth lead of a Cathode Ray Oscilloscope or the negative lead of a multimeter is connected.
On some printed circuit boards, the negative terminal of the battery (the 0v wire or terminal) is connected to a very large area of copper and this is called the EARTH PLANE or GROUND PLANE. It is designed to prevent signals travelling along the tracks (called traces) being radiated and also prevents outside interference upsetting the project. It also "tightens-up" the earth rail.

The circuit shows a 12v globe connected to a 12v battery and the circuit appears to be "broken" (not continuous).
But the **current returns** via the earth connection.

Fig 8: EARTH Return
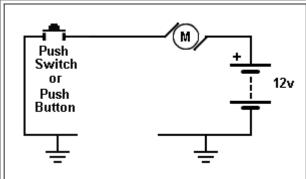
We talk about the **CURRENT RETURNING**. We don't say: the voltage returning.
The voltage of the globe must be the same as the battery voltage, otherwise it will not glow fully or it will **burn out** if it is say a 6v globe.

Fig 9: Connecting a Motor
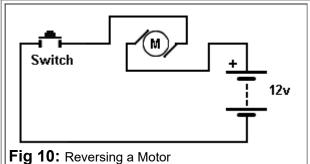
The circuit shows a 12v **Motor**.
It is turned ON when the push-switch is pressed.

Fig 10: Reversing a Motor

If the wires are connected to the motor "around the other way," the motor will reverse direction.

VOLTAGE AND CURRENT
What is voltage and what is current?
Here is a very simple description.
A battery produces a voltage called DC. (This is a very confusing name because the letter actually refer to Direct Current, so we just say **DC Voltage**).
A battery also produces current called DC - Direct Current. We say **DC current**.

VOLTAGE

Voltage is a value produced by an electrical component called a battery or cell.
A single cell produces one and a half volts. (1.5v) and although this is not a high voltage, when cells are connected together we get higher voltages.
If 6 cells are connected in series we get 9v.
Here is a 9v battery:

CURRENT

You cannot feel current with your tongue so we have to carry out another experiment:

negative terminal positive terminal

Touch the two terminals with your tongue. You get a tingle. This is a 9v tingle. Now you have "felt" electricity. This is a 9v tingle.

Place a 22 ohm or 47 ohm resistor across the terminals of the battery and hold your fingers on the resistor. It will get hot. This is the result of current flowing through the resistor and heating it up. The current will be about half an amp and the voltage is 9v, so the wattage will be about 2 to 4 watts. Feel the heat produced.

Milli - milli means 1/1,000th (one thousandth) - such as one milliamp or one millivolt.

In other words one thousand milliamps is equal to 1 amp.
One volts is not a very large value as a battery produces 9v and a cell produces 1.5v to 3.6v (depending on the type of cell.
But 1 amp is a large quality when talking about electronic circuits involving LEDs, motors and transistors.
The globe used in the experiments above requires about 300mA. (1,000mA = 1 amp)
The 3v motor used in the experiments requires about 250mA
The LEDs used in the experiments require about 20mA.
Transistors can pass about 100mA to 800mA via the collector-emitter.
In most cases current-flow in the circuits we will be discussing will be less than 1 amp and will be shown as 25mA, 100mA, 350mA etc.

WATTAGE and CAPACITY
A 9v battery has 6 very small cells and they will not last very long.
A "AAA" cell is larger and a "D" cell is much larger.

4R7

A large cell is said to have a **LARGE CAPACITY.** This means it will deliver a larger current for a longer period of time.
The **WATTAGE** of a cell is the multiplication of the voltage x current. The answer is milliwatts or watts.
The **CAPACITY** of a cell is the wattage x hours. The answer is milliwatt-hours or watt-hours. This is also called watt-hours.
You can determine the capacity of a cell (such as a rechargeable cell) by connecting it to a clock-mechanism that has a 4R7 connected across the terminals. The resistor will take a considerable current and deplete the cell in a few hours. The clock will let you know exactly how long the cell delivered the current. You can then compare other cells.
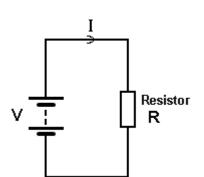
The simplest electrical circuit consists of a battery and resistor. The current flowing through the circuit will depend on the voltage of the battery and the resistance of the resistor R.

The formula connecting these three quantities is:

$$I = \frac{V}{R}$$

Ohm's Law

$$I = \frac{12}{3}$$

$$I = 4 \text{ amps}$$

This is called **Ohm's Law**. Suppose you have a 12v battery and the resistor is 3 ohms. The current flowing through the resistor will be 4 amps.
Increasing the resistance will decrease the current if the voltage remains fixed.

All the above circuits are called ELECTRICAL CIRCUITS because they contain electrical components (such as a motor, globe, relay, switch).
When the circuit contains an ELECTRONIC component such as a diode, transistor, LED, it is called an ELECTRONIC CIRCUIT or ELECTRONIC SCHEMATIC.

POWER and ENERGY

Here's an easy way to remember the difference between POWER and ENERGY:
A 9v alkaline battery has enough ENERGY to start a car. But it does not have enough POWER (strength).
Energy is effectively the strength of the battery (and this is the voltage and the current it can deliver) multiplied by the time it can deliver this energy. When the answer is obtained, it consists of three factors ((3 quantities) VOLTS, AMPS and TIME.
This results in an answer called   xxxx WATT-HOURS.
For a 9v battery the quantities are: 9 volts, 500mA and the battery will deliver this 9x0.5 = 4.5watts for about 1 hour. This is equal to 4.5 x 60 x 60 = 16,200 watt-seconds.
To start a car requires 250 amps from a 12v battery for 5 seconds.
This is: 12 x 250 x 5 = 15,000 watt-seconds.
This means the energy stored in a 9v battery could start a car if all the energy could be delivered in 5 seconds.
This is not possible however the FACT is this: A 9v battery has enough stored energy to START A CAR.

BATTERY BOOSTER

One of the simplest things we can do is start a car with a flat battery with the assistance of a BATTERY BOOSTER.
This consists of a 12v rechargeable battery in a handy case with leads to connect to the flat battery in your car.
This simple operation puts two 12v batteries in parallel, but no-one has actually described what happens and why.
That's because the explanation is very complex. We have included it here to show that a simple explanation involves a lot of technical terms and you will understand more after reading the course.

The flat battery in the car is not fully charged but it has some percentage of charge and when it sits for a period of time in a non-fully charged condition, the voltage drops from 12.6v to less than 12v as the battery gradually self-discharges due to the potential at the top of the cell being different to that at the bottom of the cell and the specific gravity of the electrolyte being different at the top and bottom. This causes an internal current to flow within the cell and slowly discharge the cell. But if you try to start the car, the voltage drops to less than 7v because the electrolyte cannot carry the high current and a slight potential is developed across the liquid.

The result is the starter-motor does not crank the car.

The reason is this: When the battery is fully charged, the current taken by the starter motor is about 300 amps. This is about 11v x 300 amps = 3300 watts = 4.4Horsepower.

But when the voltage drops to 7v, the current will drop to 190 amps to deliver 1336 watts = 1.8HP. This is only 40% of normal and that's why the car does not start. The engine needs 4HP to overcome the pressure in the cylinders due to the compression of the air during the "firing stroke."

Let's put it this way. If we have a brand new 7v battery, the car will not start. The starter-motor will only accept 190 amps when the supply is 7v.

So, we have to increase the voltage.

We do this by placing a 12v battery across the flat battery. The voltage of the flat battery will immediately rise to 12.6v. It might take 2 minutes but the flat battery will take a small current (1 to 10amps) from the battery in the "booster" and the output of the combination will be 12.6v. The current-carrying capacity of the electrolyte will improve very quickly and you have effectively given the "flat battery" a very quick charge.

The starter-motor will now accept 300 amps from the combination and **SURPRISINGLY** the cells of the "flat battery" will deliver about 200 amps and the booster battery will deliver about 100 amps. The actual sharing of current will depend on the two batteries but the secret behind the success is the increase in voltage we call **TERMINAL VOLTAGE**. The voltage on the terminals (the alligator clips).

The capacity of the booster battery is not important. It can be from 7AHr to 40AHr. We are just using a very small amount of its capacity to start the car and nearly all batteries will provide 200 Amps for a short period of time.

The voltage of the car battery is very important. The Horsepower taken by the starter-motor is defined by the formula: Pwatts = V^2/R Since the resistance remains constant, a voltage of 7 volts will produce 7x7=49 units and a voltage of 11v will produce 121units. This gives the ratio of 40% to 100% as explained above.

BATTERY BOOST

Continuing from the previous frame where we showed the effect of placing a weak battery in parallel with a good battery, we can show what happens when a weak cell is placed IN SERIES with a good cell.

This also applies when you have 5 good cells and one weak cell.

Basically, the weak cell will reduce the current. In other words, if the 5 cells are driving a motor and supplying 250mA, the 5 cells and 1 weak cell will deliver 200mA or less, depending if it is weak or very weak. The current flowing through the weak cell will have the effect of giving it a small charge - in other words, you will be charging the weak cell from the good cells when the motor is operating.

BUT . . . .

There is a way to use weak cells. If you have say 6 weak cells driving a motor and the RPM is reducing, you can add 2 more weak cells to increase the RPM.

The effect is this: The voltage from the 8 cells will be higher than from 6 cells and this will allow a higher current to flow. Sometimes the cells will provide this higher current and thus more of the energy will be delivered and you will get the last of the energy from the cells.

weak cell

good cell

INTERNAL RESISTANCE

All batteries and also all individual cells have a "secret, hidden" value of resistance inside each cell due to the resistance of the chemicals. This resistance is very small when the cell is new but it increases as the cell gets older.

It is very easy to measure this value. Simply put an ammeter directly across the cell and measure the current. Use Ohm's law to work out the resistance. But this not always a wise thing to do as some cells will deliver 10 amps and some will deliver 100 amps and damage the meter.

The diagram opposite shows a large internal resistance for the weak cell and a small internal resistance for the good cell.

If a cell did not have any INTERNAL RESISTANCE it would deliver thousands of amps. It's the Internal Resistance that limits the current.
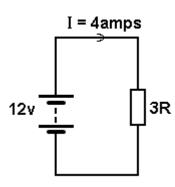
In most cases we neglect (do not consider) the value of internal resistance when making tests and when using a battery in a project.

But when a battery gets old, it cannot deliver a high current and the internal resistance gets so high that the output voltage drops from say 9v to 7v, even when the battery is not connected to a circuit.

This is the result of the INTERNAL RESISTANCE of the chemicals increasing to a point where they become noticeable and what we call "poisoning" of the chemicals due to the cell "aging" and new chemicals being produced in the cell that have a high resistance. Some of the terms we use are: "drying out and sulphating. Some cells produce spikes or needles that completely short-circuit the cell and make it totally useless.

RESISTOR WATTAGE

Resistor Wattage means two things.

1. The physical size of a resistor tells you number of watts it is capable of dissipating. This is called RESISTOR WATTAGE. It is really RESISTOR-SIZE or RESISTOR-CAPABILITY.

2. The multiplication of the voltage across a resistor and the current flowing though it will produce a value called WATTAGE. This is also called RESISTOR-WATTAGE or RESISTOR-LOSS or RESISTOR-DISSIPATION or HEAT-LOSS.
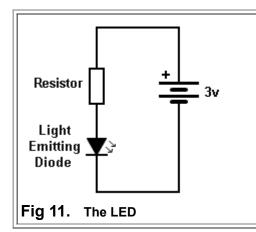
In the circuit shown, the wattage being lost in the resistor is:
12 x 4 = 48 watts.

Most of the resistors we will be using in our projects are 0.25watts. This means they will dissipate 250milliwatts, however the actual wattage being dissipated may be only 70 milliwatts and the resistor will not get hot.

0.25watts is the maximum wattage it can dissipate without overheating.

If it is dissipating 400milliwatts, it will be VERY HOT.

The wattage it is dissipating (the heat it is getting rid of) will depend on the supply voltage and the value of the surrounding components.

This resistor is dissipating 48 watts.
As a comparison, a soldering iron is dissipating about 10 to 20 watts.
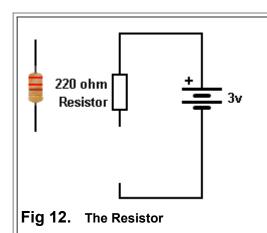
This simple **ELECTRONIC CIRCUIT** contains a **LIGHT EMITTING DIODE (LED), RESISTOR** and battery.

The circuit is classified as electronic because the LED is not an electrical item (such as a globe) but more-complex, as it produces light when current flows through a crystal and the crystal produces the colour.

Fig 11.   The LED

A **RESISTOR** must be included in the circuit to prevent the LED being damaged.
The resistor in this circuit must be 220 ohms. This is shown by the colours on the resistor. The colours for 220 ohms: **red - red - brown**. The 4th band is gold - indicating a tolerance of 5%.
A resistor has **RESISTANCE**.
It reduces the current from the battery to a required amount to prevent the LED glowing too bright.
A resistor is just like putting your foot on a hose. The water trickles out the end. The resistor "resists" the high current-flow that the battery is able to deliver.
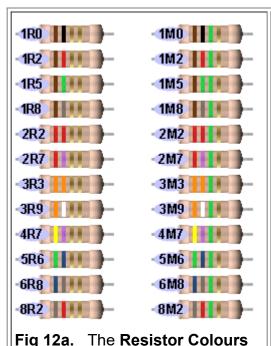
Fig 12.   **The Resistor**

There are hundreds of different resistors because the resistance-values need to cover the range one ohm to 10 million ohms.
There are also small, medium and large resistors. The resistors on the left are just a few in the range. (See the full range below). They show colour bands for 1 ohm to 8.2 ohms and 1 million ohms to 8.2 million ohms. All the other values are shown below.
An electronics engineer does not have the room to store 10 million different resistors so they make each resistor 5% or 10% higher than the previous. This reduces the number to about 100 to 200.

TOLERANCE
The first 3 bands indicate the value of the resistor and the 4th band indicates either 5% or 10% tolerance.
All modern resistors are 5% or 2% or 1%. The "old" 10% resistors are no longer made.
Gold = 5%
Silver = 10%
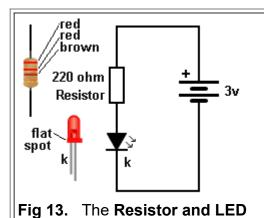
Fig 12a.   The **Resistor Colours**

The **LIGHT EMITTING DIODE** is called an electronic component (mainly because it is more complex than a globe and it produces light by a more-complex means than heating a wire).
A **LED** must be connected around the correct way. It will not illuminate if connected around the wrong way.
All **LEDs** have one lead longer than the other. The SHORT lead is called the **CATHODE** (k).
All LEDs have a flat on one side and this is the **CATHODE** lead.
The arrows on the diagram indicate light is "given off" (emitted - produced).

Fig 13.   The **Resistor and LED**
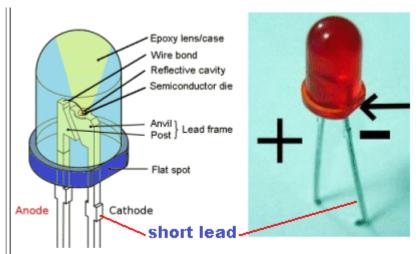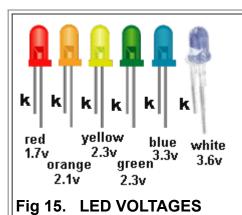
Fig 14. The LED - showing the flat spot

A close-up of a red LED. The cathode lead is the short lead and next to a flat side on the LED.
DO NOT show "+" or "-" on a diagram. Only show the letter "k" to indicate cathode.
The symbols "+" and "-" are used when a component produces a voltage or is connected directly to "+" and
"-" A LED is connected via a resistor.

to Index

Fig 15.   LED VOLTAGES

When a LED is connected to a circuit, (and the correct-value resistor is included), a voltage will be develop across the LED called the **CHARACTERISTIC VOLTAGE DROP**.
This voltage is due to the colour of the LED and the crystal inside the LED that produces the colour.
The diagram on the left shows the approximate voltage developed for each LED.
The voltage **does not change** for small, medium, surface-mount, or large LEDs.
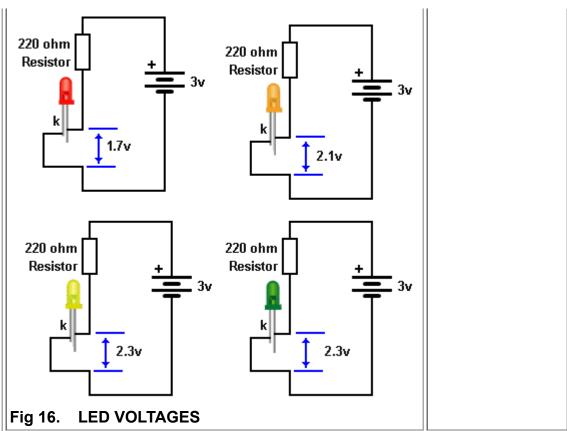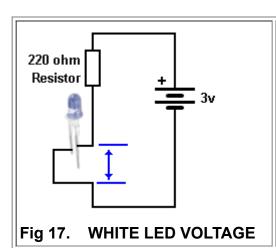
to Index

When a LED is connected to 3v battery, the following **CHARACTERISTIC VOLTAGE DROPs** will develop across each LED.
You will notice we **have not changed** the value of the resistor. It is 220R.
The LED creates the voltage and if the value of resistance is decreased, the LED will illuminate BRIGHTER.
If the LED illuminates too bright it will be DAMAGED.

Fig 16. LED VOLTAGES

Fig 17. WHITE LED VOLTAGE
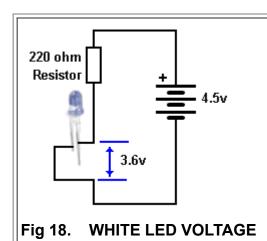
If we connect a WHITE LED to 3v supply, it will not illuminate because it needs a supply higher than 3.6v.

The resistor in series with the LED is called a **CURRENT LIMITING RESISTOR**.

In this circuit **no current flows** because the supply is not high enough.

Fig 18. WHITE LED VOLTAGE

When the supply is increased to 4.5v, the 220R resistor will allow a current to flow through the white LED and it will develop a CHARACTERISTIC VOLTAGE DROP of 3.6v across it.

The supply (the voltage of the battery) must be higher than the CHARACTERISTIC VOLTAGE DROP of the LED so the resistor will allow the correct amount of current to flow.
The ideal current for a LED is 20mA, however some LEDs will work when 1mA flows, so you have to know what you are doing.
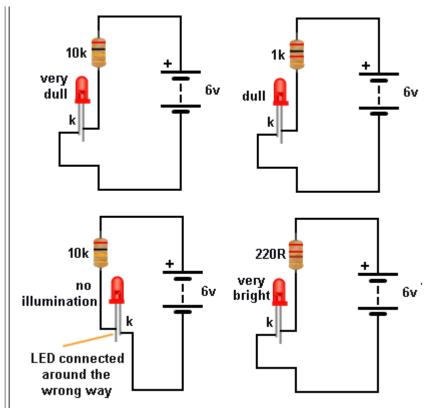
Fig 19.   Testing A LED
Now connect either the 1k, 470R or 220R and determine the brightness you need.
As the brightness increases, the current will be higher.
You can use 3v supply for all LEDs except blue and white.
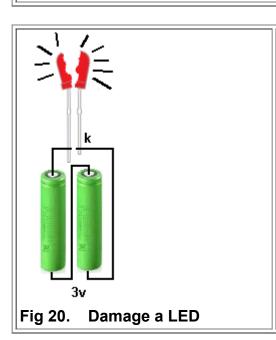
HOW TO TEST A LED
Some clear LEDs produce red or orange and some LEDs do not have the cathode lead clearly identified.
Here's how to find the colour, cathode lead and the current.
You need a 6v battery, 10k resistor, 1k resistor, 470R resistor and 220R resistor.
Connect the 6v battery and 10k resistor to the LED and it will only illuminate when the cathode is connected to the negative of the battery. This is the short lead.

Fig 20.   Damage a LED

Do not connect a 3v battery directly across a LED. It will be DAMAGED. You MUST include a resistor.

A LED IS CURRENT DRIVEN

You may have seen this statement and tried to work out what it means.

Basically it means an increase in current will make the LED brighter.

But a LED needs 2 things:

It needs a voltage that is EXACTLY the voltage required to produce illumination. And this voltage depends on the colour of the LED.

As soon as you supply the exact voltage, the crystal will begin to glow and as you increase the current, the illumination will increase.

But doing this is VERY VERY difficult.

It is very easy to supply an exact voltage such as 1.7v or 3.4v, but delivering a current such as 10mA or 20mA at the same time is very difficult. You cannot get a 1.7v battery and deliver 10mA to a LED.

As we have shown above, you need a simple components such as a resistor between the battery and LED to achieve the desired result.

A LED is CURRENT DRIVEN but firstly you need to provide a VOLTAGE that is exactly the connect value for the colour of the LED and then the current can be increased.

to Index
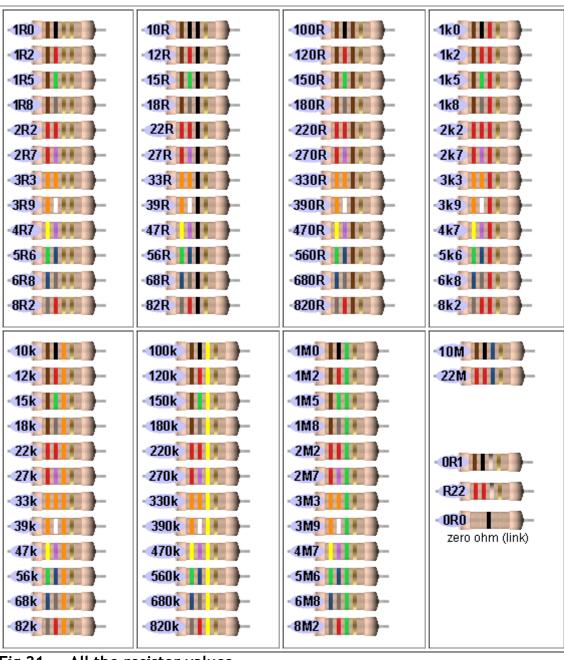
Fig 21.　All the resistor values

Here are all the colours and values for the resistors you will using in this course. Just match-up the

colours on your resistor with the resistors above and you will find the value.

Resistor values are always OHM values. One ohm is a small value. It might be the resistance of a length of wire 3 metres long.
When a switch is open the resistance is infinite - millions and millions of ohms.
The resistance of your body from one hand to the other will be about 70,000 ohms.
The resistance between two wires dipped in water will be about 1,000 to 100,000 ohms (depending on the dissolved-salts in the water - pure water has a very high resistance)
The resistance of the filament of a 3v globe will be about 30 ohms.
The resistance of the winding of a 3v motor will be about 3 ohms.
Resistors are made with values from less than one ohm to more than 10 million ohms by adding carbon to the mixture inside the resistor (and cutting a track around the outside of the resistor) then connecting a lead to each end. Adding more carbon reduces the value of resistance. Carbon has a low resistance.
Resistance-values are measured with the RESISTANCE settings on a MULTIMETER.

This is called the "Ohms Range." Sometimes with the symbol: Ω
A Multimeter will have 2, 3 4 or more scales to cover the range one ohm to 10 million ohms.
Low value resistors (from 1 ohm to 999 ohms) are written as 1R, 220R, 470R, 999R. with the

letter "R" indicating Resistance (ohms). You can also use the symbol "omega" (Ω)
For values above 1,000 ohms to 99,999 ohms, they are written as: 1k, 2k2, 4k7, 10k, 100k, 220k, 470k, with the letter "k" indicating "kilo" (thousand).
1M = 1,000,000 - one million ohms   1M2, 2M2, 4M7, 10M.
The letters "R, k and M" are placed so they take the place of the decimal point. This prevents any mistake, as a decimal point can be missing in a poor photocopy.

MULTIMETERS

There are two types of **MULTIMETER**. The top two are called **DIGITAL MULTIMETERS** (DMM) and show numbers on a display.
The lower two meters are called **ANALOGUE MULTIMETERS** and have a pointer and scale.
All meters come with a set of red and black leads.

The **red lead** is always connected to the positive of the battery or the positive on a project and the **black lead** is connected to the negative or earth or chassis.
When making a resistance measurement, the leads can be around either way.
Resistance measurements are always made with the power removed from a circuit. Any voltage on a circuit will upset the resistance reading.

to Index

Fig 22.   Resistance Measurement with Analogue Multimeter

The resistance of a resistor is measured by placing the leads of the multimeter on the ends of a resistor and turning the dial on the analogue multimeter to the resistance scale to make the pointer move to about the centre of the scale.
The resistance scale is marked with a high value on the left and 0 ohm on the right. This is opposite to all the other scales.
You must get the pointer to move to the middle of the scale as it is not accurate at left-end.
Analogue multimeters are only suitable for reading values from 1 ohm to 100,000 ohms. The scale is too hard to read above 100k.
To find the value of a resistor, you can compare the colours with the table above.

to Index

Fig 23.   Resistance Measurement with

                a DMM

A digital multimeter produces a more-accurate reading of resistance.
It is accurate from 1 ohm to 10M ohms.
Select the scale that provides a reading.

to Index

Use 47R to
make
the LEDs
bright:

220R

1.7v

1.7v

+ 6v

220R

+ 6v

1.7v
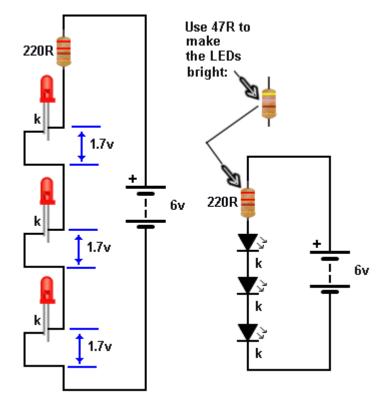
Fig 24. Connecting **LEDs in series**
LEDs can be placed in series provided the total **CHARACTERISTIC VOLTAGE DROP** across
the LEDs is LESS than the supply voltage.
In this case the voltage across the LEDs is 1.7v + 1.7v + 1.7v = 5.1v
The supply is 6v and this allows 0.9v for the CURRENT LIMITING RESISTOR.
The LEDs will not be very bright with 220R.
Change the resistor to 47R
If you connect 4 LEDs in series, the total **CHARACTERISTIC VOLTAGE DROP** will be 6.8v and
no LEDs will illuminate because the total is higher than the 6v supply.
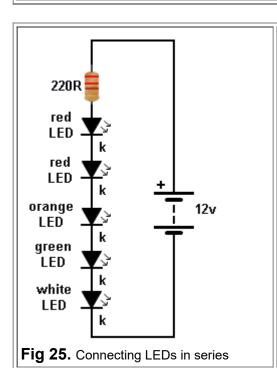
220R

red
LED

red
LED

orange
LED

+ 12v

green
LED

white
LED
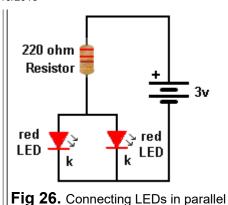
Different-colour LEDs can be connected in series.
Add up the total Characteristic Voltage for the 5
LEDs and see if it is less than 12v.

The 220R resistor will have to be reduced to 47R
to make the LEDs bright.

Fig 25. Connecting LEDs in series

LEDs can be connected in parallel if they are the
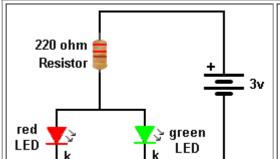
same colour.
In the diagram a red LED drops a CHARACTERISTIC VOLTAGE of 1.7v and if they are from the same manufacturer or the same batch, they will work ok.
Although we say the characteristic voltage for a red LED is 1.7v, this can change slightly from different manufacturers and one LED may glow brightly while the other is dull.
You have to build the circuit and see the result.
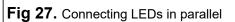
Fig 26. Connecting LEDs in parallel
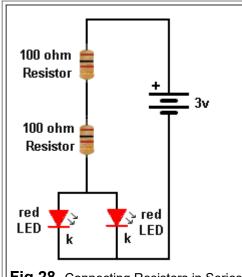
Different colour LEDs cannot be connected in parallel. The voltage across a red LED is 1.7v. This becomes the "Supply Voltage" for the green LED and it is too low. The green LED needs a supply of 2.1v to 2.3v.
Only the red LED will illuminate.

Fig 27. Connecting LEDs in parallel

Suppose you don't have a 220 ohm resistor.
You can make a 220 ohm resistor with two resistors in series. The total resistance will be 200 ohms, but resistors are not accurate and the result will be very close to 220R.
Electronic circuits are not very critical. You will not be able to see the difference in brightness between 200 ohms and 220 ohms.
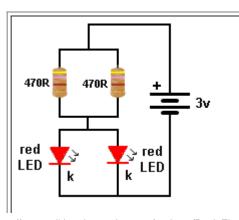
When resistors are connected in series, the total resistance is found by adding the resistance of each resistor.

$$R_{total} = R_1 + R_2 + R_3 + \ldots$$

Fig 28. Connecting Resistors in Series

You can create a 220 ohm resistor by connecting two resistors in Parallel.
When two equal-value resistors are connected in Parallel, the total resistance across the combination is HALF.
470R in parallel with 470R produces 235R.
This is very close to 220R.
We are not going into the formula as it is very complex.
Three equal-value resistors in parallel produce a total of **one-third.**
Simply get two resistors and connect them in

Fig 29. Connecting Resistors in Parallel | parallel and measure them with a multimeter.
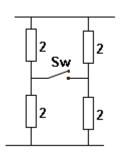
Fig 29a. Two tricky resistor questions

Figure **A** shows three resistors. It looks hard to solve so the middle resistor is turned so it connects directly to the top and bottom rail. Now you can see the circuit is three resistors in parallel. The result is one-third of an ohm.

Figure **C** shows twelve 6 ohm resistors. Replace each group with a 2 ohm resistor, because three 6 ohm resistors in parallel is equal to 2 ohms.
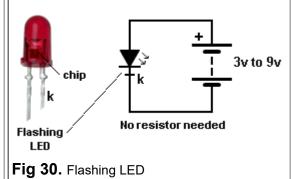
The two left resistors create 4 ohms and the two right resistors create 4 ohms. The result of two 4 ohm resistors in parallel is 2 ohms.
The resistance of the circuit does not change if the switch is open or closed.

Fig 30. Flashing LED
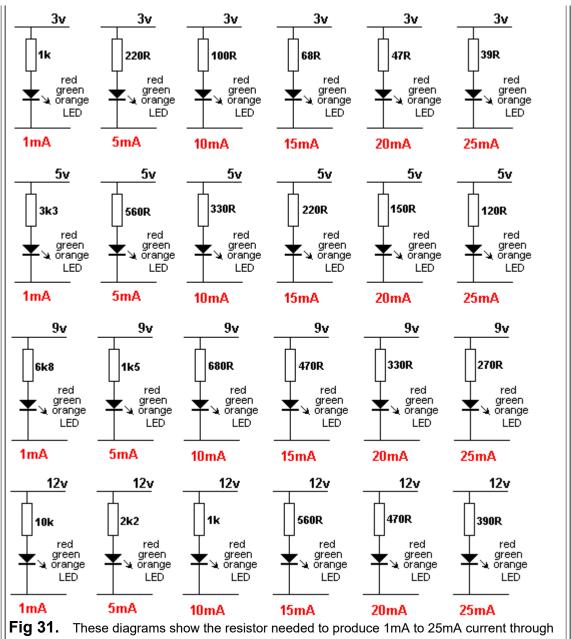
There are some special LEDs that can be connected to 3v to 9v and they flash or produce a range of colours.
These LEDs have a chip and resistor inside the body of the LED to produce the effect and allow the LED to operate on a voltage without the need for a current limiting resistor.

Fig 31. These diagrams show the resistor needed to produce 1mA to 25mA current through a single LED on 3v, 5v, 9v and 12v supply.
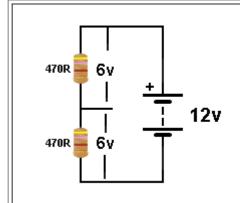
Fig 31a. Voltage Divider Circuit.

THE VOLTAGE DIVIDER

In the circuits above, the resistor and LED are forming a VOLTAGE DIVIDER.
A red LED is dropping 1.7v across it and the resistor is dropping the remaining voltage.
Whenever two (or more) components are placed across a battery, they form a **VOLTAGE DIVIDER**. Sometimes we want a 6v supply and only have 12v. We can produce the 6v supply by putting two equal-value resistors across the 12v as shown in the circuit opposite.
We are not going into the mathematics because the selection of the correct value is very complex and the circuit is very wasteful.

THE CURRENT DIVIDER

The **CURRENT DIVIDER CIRCUIT** is actually a
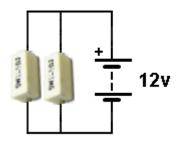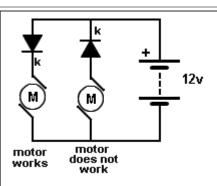
Fig 31b. Current Divider Circuit.

LOAD SHARING CIRCUIT.
Suppose you are testing a Power Supply and need a 10 watt LOAD. But you only have 5 watt resistors.
Placing two 5watt resistors in parallel across the output of the power supply will allow half the current to flow though each resistor. This is called **CURRENT SHARING** or **LOAD SHARING** and the current is divided (or passed) through each resistor according to the value of resistance.

Fig 32.   **The DIODE**

THE DIODE
The next simple electronic component is the **DIODE.**
It only works when connected correctly.
A DIODE allows current to flow through it when it is connected as shown in the diagram.
A Diode is similar to a one-way water valve.
When the diode is "facing down," the motor spins.
When it is "facing up" the motor does not spin.
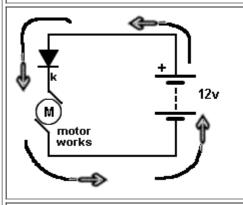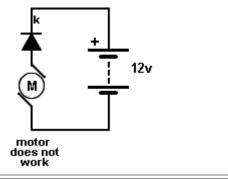
The diagram shows the "current path" around the circuit. The current is measured in AMPS and we discuss current as CONVENTIONAL CURRENT. This is the way current was thought to flow when electricity was born and they said it flows out the POSITIVE TERMINAL of the battery, around the circuit and into the NEGATIVE TERMINAL.
The arrow on the diode shows the current will flow through the diode and allow the motor to spin.
The diode is said to be **FORWARD BIASED**.

There is no flow of current because the diode prevents any current-flow when connected as shown.
The motor DOES NOT WORK.
The diode is said to be **REVERSE BIASED**.
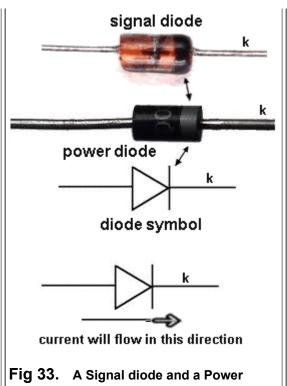
There are hundreds of different types of diodes.
Power diodes, signal diodes, low voltage diodes, high voltage diodes, high-speed diodes and many other types.
They all do one thing.
They pass current in one direction and if turned around, they **do not pass** any current.
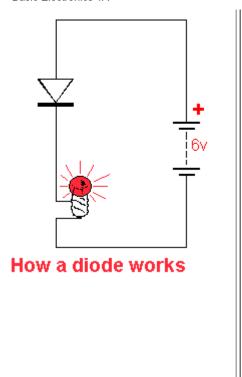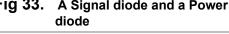
signal diode

k

power diode

k

diode symbol

k

k

current will flow in this direction

Fig 33.   **A Signal diode and a Power diode**

How a diode works

6v

current

k

protection diode

12V AMPLIFIER

12v

k   no current

protection diode

12V AMPLIFIER

12v
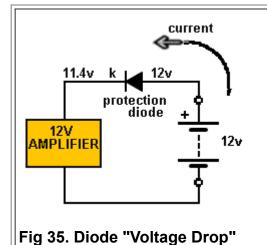
Diodes perform many function in electrical and electronic circuits. Here is an application as a **PROTECTION DIODE.**
It protects the amplifier. If the 12v battery is connected around the wrong way, no current will flow.

Fig 34. A Protection diode

current

11.4v   k   12v

protection diode

12V AMPLIFIER

12v

When a diode is placed in a circuit (and current is flowing), a small voltage develops across the diode. This voltage is called the **FORWARD VOLTAGE DROP.**
This voltage is approximately 0.6v.
This is due to a junction inside the diode where two different materials are joined.
Normally, this voltage is not important because it is only small, but sometimes you need to take it into account.
For the circuit above, the amplifier only gets 11.4v
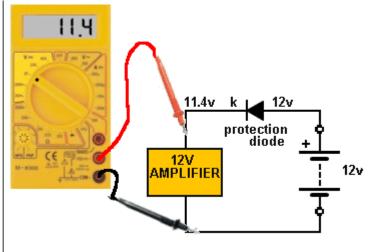
Fig 35. Diode "Voltage Drop"

Voltage is measured with a

Fig 36. Measuring Voltage with a Digital Multimeter

VOLTMETER.
Multimeters have 2 or 3 voltage ranges so you can measure low voltage (0v to 20v), medium voltages (0v to 200v and high voltages (0v to 500v).
A voltmeter is placed across the component being tested, as shown in the diagram. The Digital Multimeter is detecting 11.4v across the amplifier.

Fig 37. Measuring Voltage with a Digital Multimeter
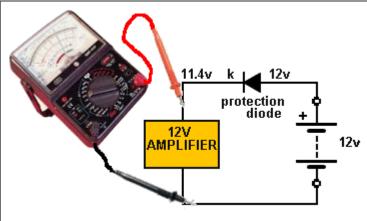
If you place the probes of a digital multimeter around the wrong way on a component, the display will show a "**-**" The meter will not be damaged.

Fig 38. Measuring Voltage with an Analogue Multimeter
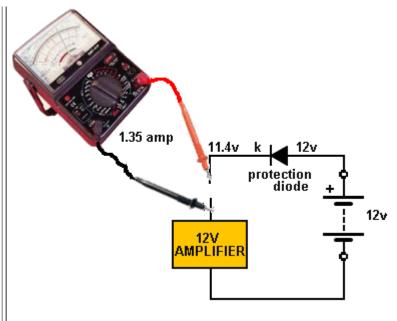
An analogue Multimeter must be connected around the correct way to make the pointer move "up scale." Select the range that will allow the pointer to show somewhere in the middle of the scale.

Current is measured by

Fig 39. Measuring Current with an Analogue Multimeter
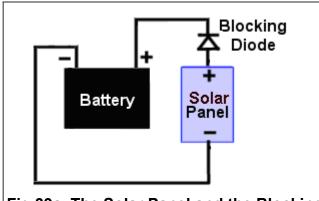
"breaking into the circuit" and inserting the leads so the positive probe is closest to the positive of the battery.
If you connect the leads around the other way, the needle will not move but it will hit the "end stop" and you may have to "bump" the meter to get the pointer to move from its jammed position.

to Index

Fig 39a. The Solar Panel and the Blocking Diode
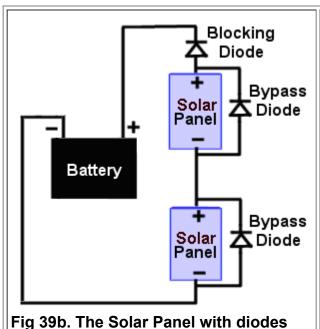
A DIODE is also used with a solar panel to prevent the battery discharging into the solar panel when the sun is not shining.
When the solar panel is not receiving any light it becomes a resistor with a large value and a small current can flow through it from the battery. The diode prevents this current-flow. The diode is called a **BLOCKING DIODE**.

to Index

Fig 39b. The Solar Panel with diodes
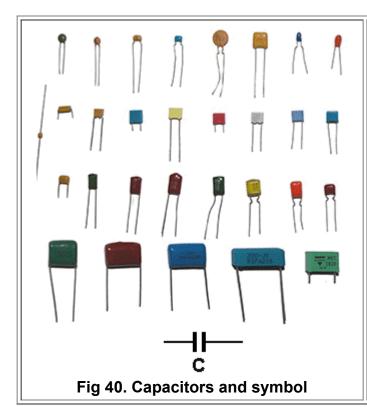
Diodes are given different names, according to their function. They all perform the same job by passing current in one direction and prevent current-flow in the opposite direction.
When the top solar panel is shaded by a cloud, it generates less current and this will reduce the current into the battery. By placing a diode across the panel, the diode will pass the current produced by the lower panel to the battery.
These diodes are called **BYPASS DIODES.**

called BLOCKING DIODES and BYPASS DIODES.

Fig 40. Capacitors and symbol

The next component we cover is the **CAPACITOR**.
There are thousands of different types of capacitor.
Each value of capacitor can have a low voltage rating, medium voltage or high voltage.
Capacitors can be very small in size and shape or very stable with temperature-rise or simply very cheap to make.
A capacitor consists of two thin sheets of metal such as aluminium with a thin sheet of plastic between. The sheets may be rolled up in a cylinder or laid on top of each other. The fact is this: the top sheet of metal does not touch the bottom sheet. This is shown in the symbol. The resistance between the two terminals is INFINITE.
The 6th capacitor in the top row is called a **MONOBLOCK**.
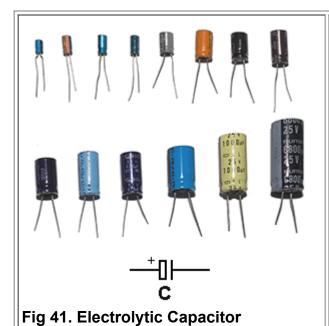
Fig 41. Electrolytic Capacitor

A capacitor gets bigger as its value increases.
It also gets bigger when the voltage-rating increases.
The basic unit of capacitance is the FARAD. A one-farad capacitor would be the size of a house. To make the capacitor smaller the sheets are etched to increase the surface-area and different insulating materials are used between the sheets.
The result is a capacitor called an **ELECTROLYTIC**. It is a bit like a rechargeable battery. It stores a lot of energy in a small space.
The negative lead is shorter and has a black stripe on the side of the electrolytic.

One FARAD is too big to handle. We use smaller values.
The middle of the range is one microfarad. This is written as 1u. (sometimes you see uF)
This is one-millionth of a FARAD.
The smallest value of capacitance is one picofarad. This is one millionth of a microfarad.
It is written as 1p.
Capacitors are broadly separated into two groups. 1p to 1u and 1u to 100,000u
Capacitors 1p to 1u are ceramic, polyester, air, styroseal, monoblock and other names.
Capacitors 1u to 100,000u are electrolytic or tantalum. A tantalum is the same as an electrolytic - for testing purposes - it is a more-compact electrolytic.

1 microfarad is one millionth of 1 farad.
1 microfarad is divided into smaller parts called nanofarad.
1,000 nanofarad = 1 microfarad
Nanofarad is divided into small parts called picofarad
1,000 picofarad = 1 nanofarad.

Recapping:
1p = 1 picofarad.   1,000p = 1n (1 nanofarad)   1,000,000p = 1u
1,000n = 1u (1 microfarad)
1,000u = 1millifarad
1,000,000u = 1 FARAD.

Examples:
All ceramic capacitors are marked in "p" (puff")
A ceramic with 22 is 22p = 22 picofarad
A ceramic with 47 is 47p = 47 picofarad
A ceramic with 470 is 470p = 470 picofarad
A ceramic with 471 is 470p = 470 picofarad
A ceramic with 101 is 100p (it can also be 100)
A ceramic with 102 is 1,000p = 1n
A ceramic with 223 is 22,000p = 22n
A ceramic with 104 is 100,000p = 100n = 0.1u   A common 100n is called a **MONOBLOCK**.
A ceramic with 105 is 1u

TYPES OF CAPACITOR
For testing purposes, there are two types of capacitor.
Capacitors from 1p to 100n are non-polar and can be inserted into a circuit around either way.
Capacitors from 1u to 100,000u are electrolytics (or tantalum) and are polarised. They must be fitted so the positive lead goes to the supply voltage and the negative lead goes to ground (or earth).

[to Index](#)

Fig 42. Charging a Capacitor

Here is an experiment to show how much (little) energy is stored in a 100u electrolytic.
When the slide-switch is in position "B", the 100u is charged by the 6v battery.
When the slide switch is moved to position "A" the electrolytic supplies energy to illuminate the red LED via the 220R resistor. It will illuminate for a short period of time.
By moving the switch back and forth, you can keep the LED illuminated.
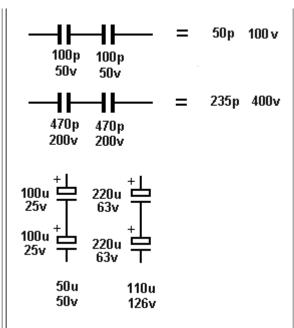
[to Index](#)

Capacitors can be connected in **Series or Parallel** to obtain a value of capacitance you may not have available.
They are also connected in series to increase the effective **VOLTAGE RATING**.
However when two equal-value capacitors are connected in series, the final value is HALF, and thus you need two with double the final-value to get a value with an increased voltage-rating.
When two equal-value capacitors are connected in series, the result is HALF.
(This is the opposite to connecting resistors)
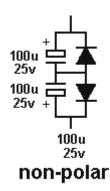
Fig 43. Capacitors in Series

Fig 43a. Non-polar Capacitor

Non-polar Capacitor (electrolytic)

A normal electrolytic must be connected the correct way in a circuit because it has a thin insulating layer covering the plates that has a high resistance.

If you connect the electrolytic around the wrong way, this layer "breaks-down" and the resistance of the electrolytic becomes very small and a high current flows. This heats up the electrolytic and the current increases. Very soon the capacitor produces gasses and explodes.

One big mistake in many text books shows how to make a non-polar electrolytic by connecting two "back-to-back."

They claim 2 x 100u connected back-to-back is equal to 47u.

This appears to be case when testing on a meter but the meter simply charges them for a short period of time to get a reading.

If you allow them to charge fully you will find the reverse electrolytic has a very small voltage across it.

Secondly, when you are charging them, you are putting a high current through the reverse electrolytic and damaging the layer.

To prevent this, you need to add two diodes as shown in the diagram.

In addition, 2 x 100u "back-to-back" is very near 100u.

Capacitors can be connected in **Parallel** to obtain a value of capacitance you may not have available. (This does not change the **VOLTAGE RATING**.) When two equal-value capacitors are connected in parallel, the result is DOUBLE. (This is the opposite to connecting resistors). If one electrolytic is 25v

and the other 63v, the answer is the LOWER VOLTAGE = 25v.

$$C_{total} = C_1 + C_2 + C_3 + \cdots\cdots$$

Fig 44. Capacitors in Parallel

Fig 45. Variable Capacitors and Resistors
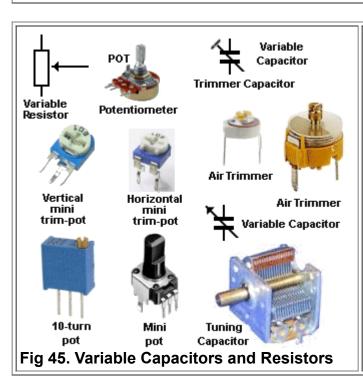
The value of a capacitor or resistor may need to be increased or decreased in a circuit to tune in radio stations or increase and decrease the volume of a speaker. The symbol for these components have an arrow to show they can be adjusted.
The resistance of a potentiometer can be from 1 ohm to 5M
They come in many different shapes and sizes to suit the PC board or front-panel layout.
The "T" represents a trimmer capacitor and this can be from 1p to about 120p.
A variable capacitor will be from about 10p to 415p.
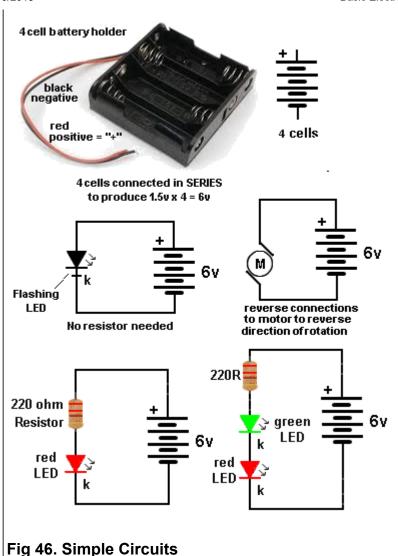
Simple CIRCUITS
We have covered enough symbols and components to create a number of simple circuits.
These circuits will show how to connect a motor, a LED, (how to make it bright or dull) and how to connect 4 cells to make a battery.

Note: The flashing LED

4 cell battery holder

black negative

red positive = "+"

4 cells connected in SERIES to produce 1.5v x 4 = 6v

4 cells

6v

Flashing LED

No resistor needed

reverse connections to motor to reverse direction of rotation

6v

220 ohm Resistor

red LED

k

6v

220R

green LED

k

red LED

k

6v

does not need a resistor because a resistor and chip are inside the LED, to make it flash and control the current.

Connect all the components around the correct way and then connect them around the wrong way to see what happens.
Connect the flashing LED in series with a red LED and see what happens.

Fig 46. Simple Circuits

QUESTIONS
1. Explain why the Flashing LED circuit has no external resistor.
2. How many 1.5v cells are needed to produce a 6v battery
3. Explain what happens when you reverse the leads to a motor.
4. Identify the positive terminal:

5. Can 3 green LEDs be connected in series to a 6v supply?
6. A variable resistor is also called: _____
7. The combined resistance of two 1k resistors in series is: _____
8. The combined resistance of two 1k resistors in parallel is: _____
9. Name the short lead on a LED
10. Name the type of multimeter with a pointer and scale: _____
11. The total capacitance of two 100u electrolytics in series is: _____
12. The total capacitance of two 100u electrolytics in series is: _____
13. Write these values in words:
      22R   _____
      4k7   _____
      1n   _____
      100u _____
14. How many 1.5v cells in a 9v battery?

15. The red probe is: _____   (positive/negative)

16. Conventional current flows from: _____   (positive to negative / negative to positive)
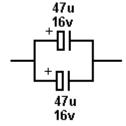
17. Make a 470u electrolytic with two electrolytics:

18. What is the voltage drop across a diode?

19. Name the component that only allows current to flow in one direction: _____

20. Name this symbol:

21. When resistors are connected in series, the resistance of the combination: _____
(increases / decreases)
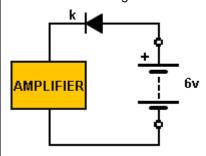
22. When two capacitors are connected in parallel, the voltage-rating of the combination:
_____ (increases / equal to the capacitor with the lowest voltage-rating)

23. Draw two 2k2 resistors in parallel.

24. Which is larger: 470R or 22k

25. What is the value of this combination:
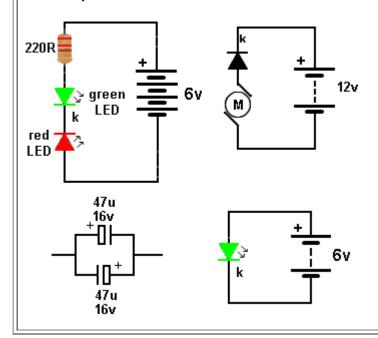
26. What is the name of the resistor in series with a LED: _____

27. What is the voltage across the amplifier: _____
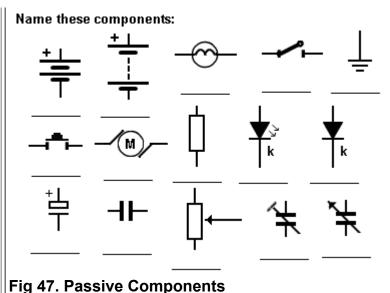
28. Identify the fault with these circuits:
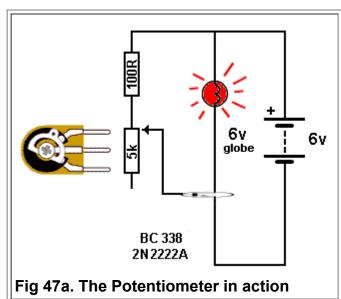
Passive Components
All the components shown

Name these components:

on the left are called PASSIVE COMPONENTS. This means they do not amplify.
Write the name beside each symbol.

Fig 47. Passive Components

to Index

The Potentiometer is a variable resistor.
It consists of a curved carbon track with a wiper that touches the track and can be turned via a screwdriver or knob.
The wiper is the middle wire on the circuit symbol and it moves up and down as shown in the animation.
When the three leads are connected the symbol is called a potentiometer.
When two leads are connected it is a variable resistor.
When the resistance increases, less current flows through the pot.

Fig 47a. The Potentiometer in action

to Index

variable resistor   potentiometer

There is a difference in operation between a **Variable Resistor** and a **Potentiometer**.
Both will increase or decrease the sound level as a volume control or the speed of a motor or the brightness of a globe, but a Potentiometer will guarantee zero volume or zero brightness when the pot is turned fully anticlockwise (as shown in the animation).
This is because the output will be zero volts, but the variable resistor may still deliver some "energy" (voltage and current) to the circuit when turned fully anticlockwise.

Fig 47b. The Variable Resistor and Potentiometer

to Index

Potentiometers come in values from 100 ohms to 5 Meg ohms (500R, 1k, 2k, 5k, 10k, 50k, 100k, 200k, 250k, 500k, 1M are most popular).
They come as linear, or logarithmic where the resistance of the track (per mm) is higher at one end. Because our hearing is not linear, these pots can be used as volume controls to produce a gradual (very nearly linear) increase in volume.
Selecting the correct value of resistance for a circuit is VERY complex. If the value is not correct, the volume will not be loud or it will drop to zero before the pot is turned fully anticlockwise. Or
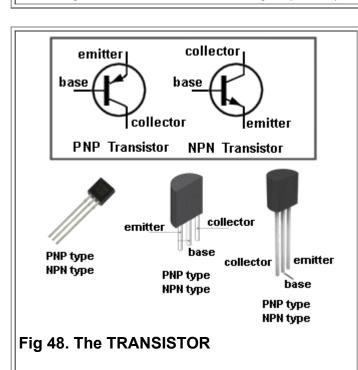
the motor will drop to zero at mid-turn of the pot or it will not reduce in RPM to the desired amount.
The simple answer is to copy a circuit.
Or you can try the whole range of pots and you will find one value is the best.
A Potentiometer can be used in hundreds of different circuits to produce hundreds of different effects, but the actual "thing" that flows between the input and output is a percentage of the voltage. At the same time the current will also be passed to the output at a reduced value. A pot actually delivers BOTH reduced values at the same time and the receiving circuit will be designed to "look for" the change in voltage or current. If the supply voltage is not rising or falling, the "values" are called DC values.
The voltage can also be in the form of a signal (volume). This is called an AC signal.

The TRANSISTOR

Fig 48. The TRANSISTOR

A TRANSISTOR is an ACTIVE device. It AMPLIFIES.
There are many types of transistor (over 20,000 different types) from hundreds of manufacturers and they have many different names.
We are going to study the simplest. It has the technical name **BIPOLAR JUNCTION TRANSISTOR (BJT)** but we are going to call it a **TRANSISTOR**.
There are two types in this group: **PNP** and **NPN**.

The type we will study is also called a **SMALL-SIGNAL TRANSISTOR.**

You cannot tell an **NPN** transistor from **PNP** by looking at it. You must test it in a circuit.
In **Fig 65** you will make a **Transistor Tester** project, but first some basic facts:

Fig 49. The NPN TRANSISTOR in a Circuit

The first type of transistor we are going to study is the **NPN**.
A transistor has three leads: **BASE, COLLECTOR** and **EMITTER**
Basically, a small current enters the base and a large current flows through the collector-emitter leads as shown in the diagram.
The resistor in the collector lead is called the LOAD Resistor.
Sometimes the load is a speaker.
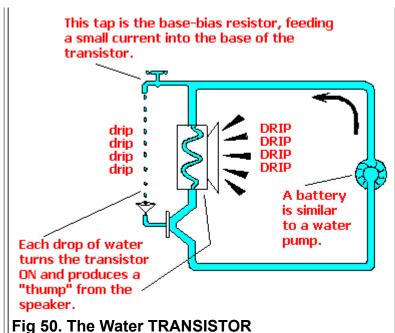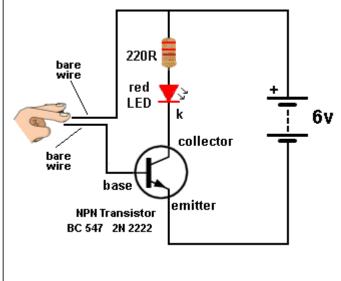
The transistor is similar to the diagram opposite. A small drop of water entering the base is amplified to

This tap is the base-bias resistor, feeding a small current into the base of the transistor.

drip
drip
drip
drip

DRIP
DRIP
DRIP
DRIP

A battery is similar to a water pump.

Each drop of water turns the transistor ON and produces a "thump" from the speaker.

produce a loud DRIP from the speaker.

Fig 50. The Water TRANSISTOR

bare wire

bare wire

220R

red LED

k

collector

base

emitter

NPN Transistor BC 547 2N 2222

+

6v

In this experiment we will construct a ONE TRANSISTOR circuit similar to the WATER TRANSISTOR above and observe the results.
Make sure the two leads DO NOT TOUCH. If they touch, the transistor will be DESTROYED.
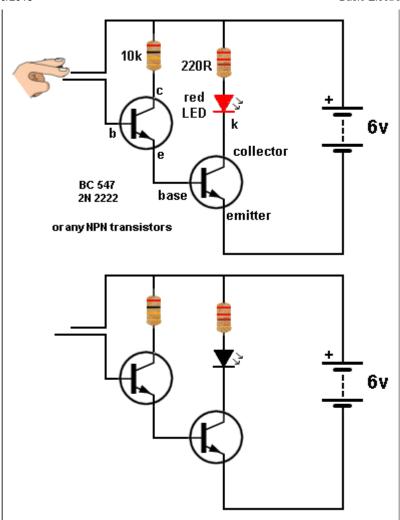
The transistor is amplifying the current through your finger via the two leads and it will be very dim.

ANIMATION

The lower diagram shows the transistor turning ON when a finger is pressed against the two wires. The finger produces a resistance that turns the transistor ON and this turns the transistor into a smaller and smaller resistor. That's how more and more current flows through the LED and it gets brighter and brighter.

+

6v

Fig 51. One Transistor Circuit

By adding another transistor we amplify the

Fig 52. Two Transistor Circuit

current through the finger about 200 times and now the LED will glow bright.

Make sure the bare wires do not touch each other as this will destroy BOTH transistors.

ANIMATION
The lower diagram shows both transistors turning **ON** when a finger is pressed against the two wires.
They both becomes smaller and smaller resistors.
The first transistor allows more current to flow into the base of the second transistor and this is how the second transistor turns on more and more. This allows more current to flow through the LED and it gets brighter and brighter.
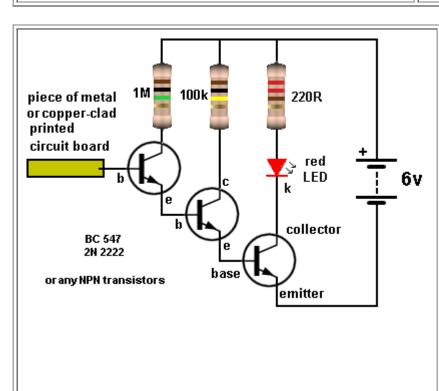
to Index

This circuit has enormous gain. Each transistor has a gain or more than 200 and the final gain will be more than:
200 x 200 x 200 = 8,000,000
8 MILLION!

The circuit is very sensitive to static voltages in the air or electrical waves such as the waveform produced by the electrical wiring in a house.
Move the project around a room and detect all the electrical signals.

Fig 53. Three Transistors
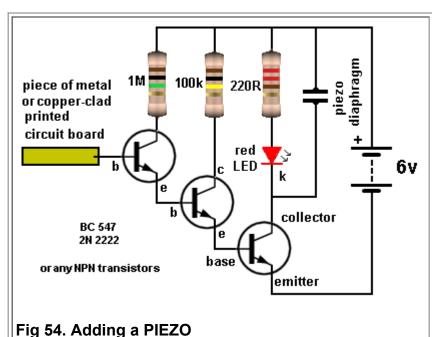
You can see the effect of one transistor. It does not do much.
The two transistor circuit allows the resistance of your finger to deliver current into the base of the first transistor and this transistor delivers more current into the base of the second transistor.
The result is more collector-emitter current and the LED illuminates.
The three transistor circuit produce an ENORMOUS effect.
It will pick up STATIC ELECTRICITY and all forms of electro-magnetic energy (radiation) and illuminate the LED.

Fig 54. Adding a PIEZO

By adding a piezo diaphragm to the output you will be able to hear the hum of the mains.
This is the frequency of the supply into your house. It will be either 50 cycles per second or 60 cycles per second.
The term: "cycles per second" is given the name HERTZ after **Heinrich Rudolf Hertz**, who was the first to prove the existence of electromagnetic waves.
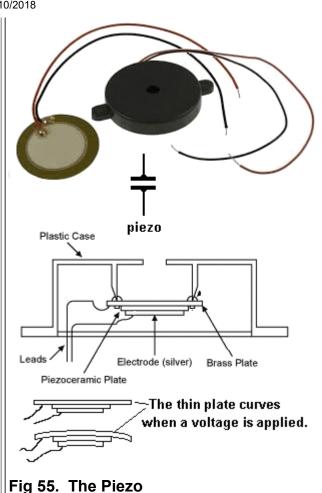
The Piezo diaphragm is held around the outer edge inside a plastic case and when a voltage is applied to the two leads, the thin plate curves very slightly.
When the voltage is removed, the plate returns to its flat shape.
If the voltage is reversed, the plate curves in the opposite direction.

The curving is due to a thin layer of ceramic material under the plate and
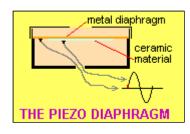
then a film of metal is deposited onto the ceramic so a lead can be soldered.
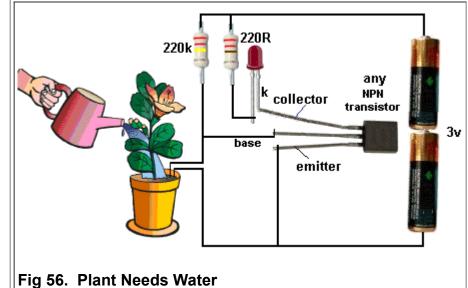
There is infinite resistance between the two leads as the ceramic material is an INSULATOR.
The capacitance between the two leads is approx 22n.
The Piezo is a passive device. It needs a pulse or frequency for it to produce an output.

THE PIEZO DIAPHRAGM

Fig 55. The Piezo

The **ONE TRANSISTOR CIRCUIT** above can be turned into a detector to show when a plant needs water.
Place the two probes into the soil and water the plant. The LED will turn off. As the water evaporates the LED will turn **ON** to let you know the plant needs watering.

Fig 56. Plant Needs Water
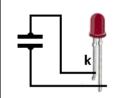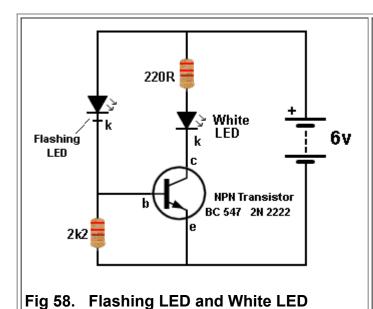
This experiment produces a pulse from the piezo when it is tapped and the LED illuminates briefly.

The LED can be connected either way around.

This proves the diaphragm flexes when a voltage is applied and also in the reverse situation. A voltage is produces when the diaphragm is tapped.

Fig 57. Tap the Piezo

A flashing LED is not very bright.
It can be connected to a transistor and the transistor will drive a very bright white LED.
The transistor is an amplifier.
It is amplifying the current flowing through the flashing LED and supplying a higher current for the white LED.

We cannot discuss any further details of the circuit at the moment because the actual operation of the circuit is quite complex.

At the moment we just need to experiment with simple transistor circuits.

Fig 58.   Flashing LED and White LED

Fig 59. Soldering Iron
We now come to the point of HELPING YOU WITH CONSTRUCTION.
We have already shown you 6 different circuits and there are many ways to build them.
You can:
1. Solder them.
2. Build them on an Experimenter Board
3. Connect the components with clips or twist the leads together.
It does not matter how you build the circuits.
The fact is this: YOU MUST START BUILDING.
The best soldering iron for a beginner is a CONSTANT TEMPERATURE soldering iron.
It has a dial that can be turned to set the desired temperature.
An ordinary soldering iron GETS TOO HOT. It is not suitable for soldering electronic circuits.
This is something that no-one has mentioned before. An ordinary soldering iron will melt the solder TOO QUICKLY and burn the resin inside the solder and make soldering very difficult for a beginner.
Soldering must be done slowly so the resin in the middle of the solder gets hot and cleans the leads of the components so the solder will "stick."
That's why you must apply the solder to the leads you are soldering and allow the resin to "attack" the leads and clean them.
The cheapest TEMPERATURE CONTROLLED soldering Iron is available on eBay for les than $10.00 (post FREE).
You will also need a small roll of solder (0.9mm) and a soldering Iron stand.
Email Colin Mitchell for links to eBay. (talking@tpg.com.au)
A whole book could be written on the ART OF SOLDERING.
Look on the web for articles and videos on SOLDERING.

SOLDERING

1. Soldering is very easy and very simple. You just need a **Temperature Controlled Soldering Iron**, **Fine solder** and clean components.

2. Remember this: It is **NOT** the solder you need for a joint, but the **FLUX**. And the flux lasts for only 2 seconds. When the flux is **HOT** it attacks and cleans the joint so that the solder will stick.

3. Turn ON the Temperature Controlled Soldering Iron to a low temperature. Put solder on the tip. It will not melt. Turn up the temperature slightly. Try more solder. As soon as the solder starts to melt, this is your starting point. Turn up the temperature slightly MORE and this is the correct temperature for **small, delicate, fine soldering**.

4. Place a component through a hole and bend the lead slightly so the component does not move. Turn the board over and touch the iron on the component and bring the solder **FROM THE OTHER SIDE** so the solder melts and flows towards the iron.

From start-to-finish, count one-two-three and remove the solder. Count four-five and remove the iron. You will have a perfect joint.

If you are soldering thick leads or large pads on a circuit board, you will need to turn the temperature UP slightly.
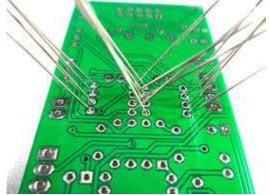
You must add enough solder to make the joint "bulge" slightly.

Fine solder (1mm or 0.9mm or 0.8mm) makes the best joint because it is easier to use.

Use a wet sponge to clean the tip or a ball of "Steel Wool." Steel wool is the best.
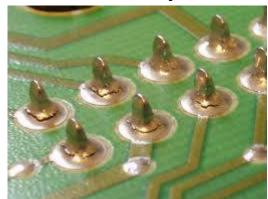
Here is the steel wool, bending the leads and some examples of poor joints due to insufficient solder:

Steel wool cleans the tip beautifully

Bend the leads before soldering

The joints do not have enough solder and that's why they fractured. Called a DRY JOINT.
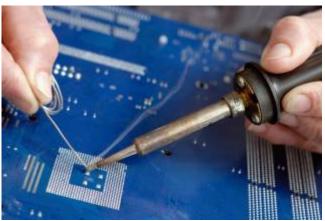
More "Dry Joints."

This is the cheapest and simplest soldering iron stand.

This stand is very messy as the spring grabs the iron and makes it difficult to remove from the stand. Test the stand before buying. You will se why not to buy this type of soldering stand. Get one with a "wide mouth" and a heavy stand is best as it does not move.

This photo clearly shows how to hold a soldering iron and solder.

This is NOT a temperature-controlled soldering iron and you can see it is too hot as it is burning off the flux too quickly.
Temperature Controlled Soldering Irons are now cheaper than the JUNK soldering iron shown in the photo. See eBay for prices.

BREADBOARD
The term BREADBOARD refers to any piece of wood or plastic containing pins or pegs or clips or holes where you can build a circuit.
The components can be soldered, twisted clipped or fitted into holes.
Breadboard also means the circuit can be easily pulled apart.
Some breadboards do not have two rows for the positive and negative rails. Connections under the board for the positive rail is shown with a black line in the photo. Connections on the main section of the board are shown with blue lines. Your breadboard MUST look exactly like the photo opposite.
Other breadboards are quite useless.
The breadboard in the photo can be purchased on eBay for less than $5.00 (post FREE).

Photo shows a number of components fitted to the breadboard.
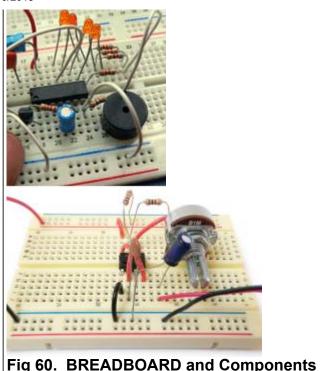
Fig 60. BREADBOARD and Components

to Index

Fig 61. Jumpers

The components on the BREADBOARD are fitted down the holes and metal strips under the board join each column of 5 holes. If you want to join one hole with another, you can use 0.5mm tinned copper wire or JUMPERS. See photo opposite. Jumpers can be purchased on eBay for less than $3.00 posted.

Email Colin Mitchell for links to eBay. (talking@tpg.com.au)

to Index
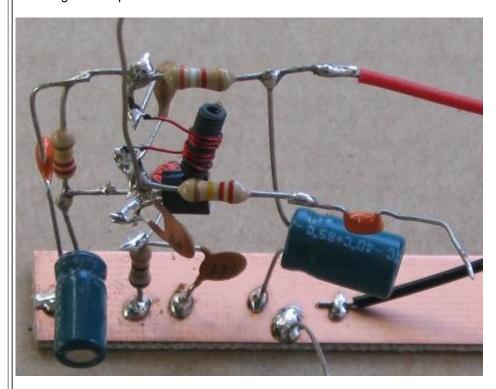
Fig 62. Breadboard with Nails

If you don't have a soldering iron or experimental breadboard, you can make your own board with nails. See the photo above. It is a multivibrator circuit and we will be presenting this circuit in a moment. The components can be twisted around the nails and bare wire used to join some of the nails to complete the circuit.

Another method of connecting the components is called BIRD-NESTING. This involves soldering the components "in the air" as shown in the 27MHz transmitter circuit below:

Another way to connect the component(if you don't have a soldering iron), is to wind 6 turns of bare wire around each connection and leaving all the components "in the air." The bare wire can be obtained from hook-up flex. This is plastic coated "wire" containing up to 15 fine strands of wire. Use a single strand for the connections. None of the components will touch each other BY MISTAKE and the circuit will work perfectly. Bird-nesting is a good way to build a quick circuit and test its performance. It might look messy but you can easily change any component.

to Index
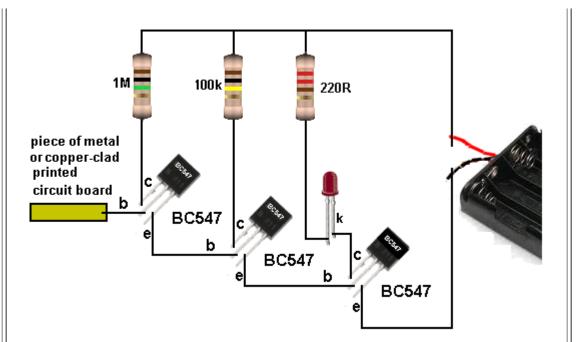
Fig 63. Wiring the 3-Transistor Circuit using BC547 Transistors
The diagram shows how to connect 3 x BC 547 transistors.
The leads of a transistor can be collector-base-emitter OR
emitter-base-collector and that's why we have provided 2 different wiring diagrams.

to Index
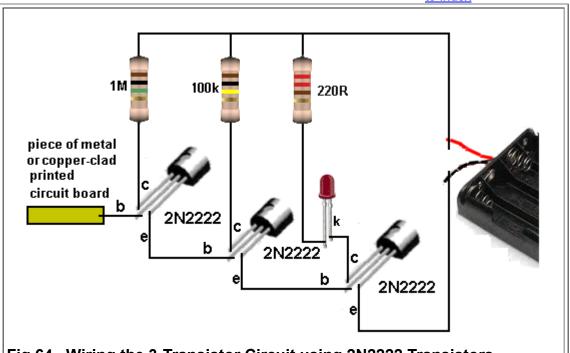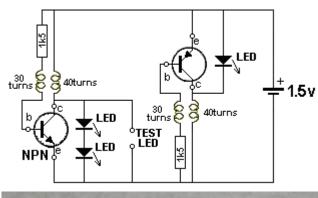
Fig 64. Wiring the 3-Transistor Circuit using 2N2222 Transistors
The diagram shows how to connect 3 x 2N2222 transistors.

to Index
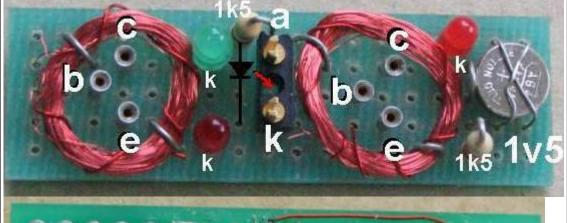
Fig 65. Transistor Tester
This handy transistor and LED tester can be built to test LEDs and both PNP and NPN transistors.
The project consists of two identical circuits, one for NPN and one for PNP.
You can build just the NPN section and then build the PNP section later.

LED VOLTAGE
We have shown a LED needs at least 1.7v supply for it to operate.
This circuit works on 1.5v and thus the action of the transistor and coil (called a Transformer) MUST be increasing the voltage for the LED to illuminate.
This circuit works on two "actions."
1. Transistor ACTION - this is the action of a transistor providing gain to make the circuit oscillate.
2. Transformer ACTION - this is the action of a coil of wire producing a voltage higher than the

supply voltage when it is turned off.

This circuit is very technical and very complex.
We will be explaining it in a very simple way because this is a **Basic Electronics Course**.

THE TRANSFORMER - the two coils of wire on the left and the two coils of wire on the right.
When the voltage (actually the current) is switched off, the 40 turn coil in either of the circuits in this project; the voltage across the coil rises to more than the 1.5v supply and is in the opposite direction to the voltage of the supply.
The circuit looks to be very simple but it uses an air-cored transformer to produce the voltage needed to illuminate the LED indicators and the circuit only works when the transistor is connected correctly. There are two separate circuits, one for NPN transistors and one for PNP transistors. We will cover the NPN section:
The circuit turns ON when the NPN transistor is fitted and the current through the 30 turn coil and 1k5 resistor turns ON the transistor and produces expanding flux in the 40 turn coil. This flux cuts the turns of the 30 turn coil and produces a voltage in the coil that adds to the supply voltage and increases the current into the base. This turns the NPN transistor ON more. This action continues until the transistor is fully turned ON. At this point the current in the 40 turn coil is a maximum but it is not expanding flux and the 30 turn coil ceases to see the extra voltage. Thus the current into the base reduces and this turns the transistor OFF slightly. The flux produced by the 40 turn coil now becomes collapsing (or reducing) flux and it produces a voltage in the opposite direction to greatly reduce the current into the base. In a very short period of time the transistor becomes TURNED OFF and it is effectively removed from the circuit. The flux in the 40 turn coil collapses quickly and it produces a voltage in the 40 turn coil that is higher than the supply voltage and is in the opposite direction. This means the voltage produced by the 40 turns ADDS to the supply voltage and is delivered to the LEDs to illuminate them.
The NPN circuit has two LEDs in series so that a LED of any colour (including white) can be connected to the TEST LED terminals and it will illuminate. You can use any colour LED for any of the LEDs, however it is best to use either green or yellow or white for the single LED.
The two "coils" are wound on a 10mm dia pen with 0.1mm wire (very fine wire). The loops of tinned copper wire holding the coils on the board are connected to separate lands under the board and MUST NOT produce a complete loop as this will create a "Shorted Turn" and the circuit WILL NOT WORK.
If the LEDs do not illuminate, simply reverse the wires to the 30 turn coil.
The circuit does not need an ON/OFF switch because the LEDs require a voltage of over 2v to illuminate (the orange LED) and the supply is only 1.5v. A red LED needs about 1.5v to 1.7v to operate but when it is in series with a green LED, this voltage is over 3.5v.
All the components fit on a small matrix board 5 holes x 18 holes. A kit of parts for the project is available for $4.00 plus $3.00 postage and ordering details can be obtained by emailing Colin Mitchell. (talking@tpg.com.au)
Build the circuit and test your transistors and LEDs.
We will be covering more on the action of a transistor and the action of a transformer in the discussion below, but it is important to build the circuit and see it working.
It is your first piece of **TEST EQUIPMENT**.

Questions
1. Identify the letters "c" "b" and "e"
2. What type of transistor is tested in the first set of hollow pins?
3. Put a PNP transistor into the first set of hollow pins and try all positions. Does the red and green LEDs illuminate?
4. When both the red and green LEDs illuminate, what is the approximate voltage across the pair?
5. When you fit a red LED to the test-socket, what is the approximate voltage across it?
6. When you fit a red LED to the test-socket, why does the red LED and green LED on the PC board turn off?
7. Why doesn't the project need an on/off switch?
8. The two coils for the circuit on the left is called a TRANSFORMER. Do the connections of the windings have to be connected to the circuit around a particular way?

ROBOT MAN
This multivibrator circuit will flash the Robot Man's eyes as shown
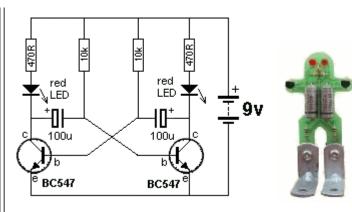
Fig 66. ROBOT MAN
The ASTABLE MULTIVIBRATOR or "free-running" multivibrator.
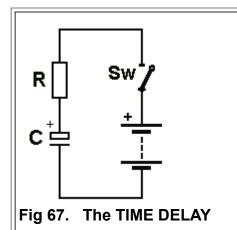
in the photo. The kit of components is available from Talking Electronics for $8.50 plus postage. Send an email to find out the cost of postage: talking@tpg.com.au
The photo shows the LEDs flashing.
The circuit is called an **ASTABLE MULTIVIBRATOR** and this means it is not stable but keeps switching from one transistor to the other.
It is also called a **FLIP FLOP** circuit.

to Index

Fig 67.   The TIME DELAY

The **TIME DELAY** circuit consists of a Resistor **R** and Capacitor **C** in **SERIES**.
When the switch is closed, the electrolytic (called the CAPACITOR) charges slowly because the resistor only allows a small amount of current to flow.
It's just like charging your mobile phone. The battery takes time to charge because there is a resistor in the circuit to limit the current. If we remove the resistor in the mobile phone, the battery will get too hot when it is being charged but in the **TIME DELAY** circuit, we want the capacitor to charge slowly, because we want a **TIME DELAY.** .

to Index

Fig 68. The charging of a capacitor is the same as building a brick wall.

CHARGING A CAPACITOR
The capacitor in **Fig 67** charges via the resistor **R**. But the voltage on the capacitor does not rise at a constant rate.
It starts off charging very quickly and as the voltage across it get higher, the voltage increases at a slower and slower rate.
In the photo I am building a brick wall.
I am working at a constant rate.
When I started building the brick wall, I laid 5 rows of bricks (5 courses) in the first hour.
As the wall increased in height, I had to climb the ladder and I could only lay 3 courses an hour and finally the wall was so high I could only lay 1 course per hour.
This is exactly the same as a capacitor charging. When the capacitor is uncharged, the supply voltage allows a high current to pass through the resistor **R** and the energy quickly fills the capacitor. This results in a rapidly increasing voltage on the capacitor. But as the voltage on the capacitor increases, the difference in voltage between that on the capacitor and the supply is very small and only a small current will pass through the resistor. This means the voltage on the capacitor increases at a slower rate.
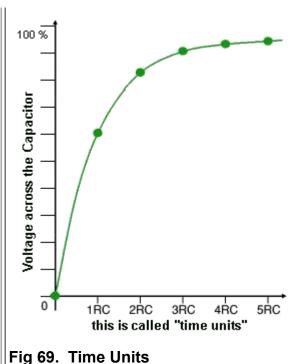
to Index

Fig 69. Time Units
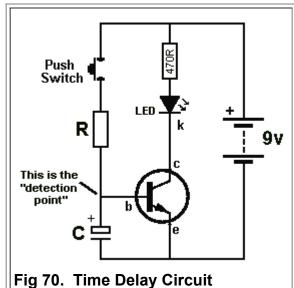
It really does not matter how fast or slow or uneven a capacitor charges because most circuits detect a voltage on a capacitor and the time taken to reach this voltage is called the **TIME DELAY**.
But to prevent you thinking the capacitor charges "smoothly" we have to explain what actually happens.
The graph on the left shows the capacitor charging. You can see it charges quickly at the beginning and then charges slowly and then very slowly.
You can see the first part of the graph is fairly "straight" (constant charging) - NOT "straight up and down" but a straight line - and this applies to a voltage of about 63%.
The time taken to reach this voltage is called **ONE TIME UNIT** - also called **ONE TIME CONSTANT**. The graph continues for another 4 "time units" (time-constants) and the final voltage is very nearly 100%.   (It never reaches 100%.)

to Index

Fig 70. Time Delay Circuit
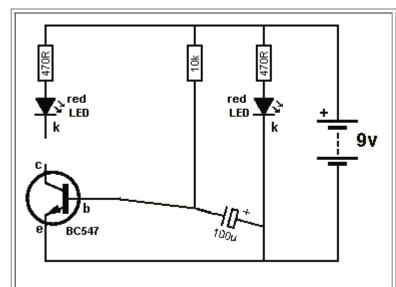
A **TIME DELAY** circuit needs three things:
1. A Resistor (R)
2. A Capacitor (C)
3. A "Detection Point."

When the switch is pressed, capacitor (**C**) takes time to charge via resistor (**R**) and after a short period of time the voltage at the **DETECTION POINT** is 0.6v and the transistor is TURNED ON. The LED illuminates.

Build the circuit with 100u and 100k and see how long it takes before the LED illuminates.

to Index

In the **ROBOT MAN** project you can see the "**TIME DELAY**" circuit made up of the 100u, 10k resistor and the base of the transistor.

This is one of the most important **BUILDING BLOCKS** in electronics.
It is the basis of all oscillators and will be discussed below, after we explain a few more details.

**Fig 71. The "TIME DELAY" in the ROBOT MAN
Project**

Fig 72. Turning A Transistor ON

We will use the **TIME DELAY** circuit to turn the transistor **ON**.
Make sure the 100u is uncharged by touching both leads (both ends of the capacitor) at the same time with a JUMPER - this is a piece of wire shown in Fig 61.
Push the switch and noting happens. After a short period of time the LED starts to glow and then comes on fully.
This shows two things:
1. The transistor is not turned ON when the base voltage is zero.
2. The base voltage must be 0.6v for the transistor to start to turn ON and when the voltage is 0.65v the transistor is turned **ON FULLY.**
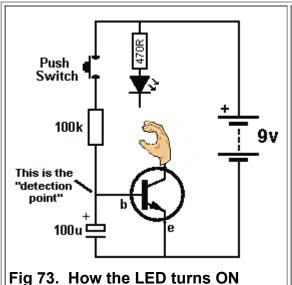
Fig 73. How the LED turns ON
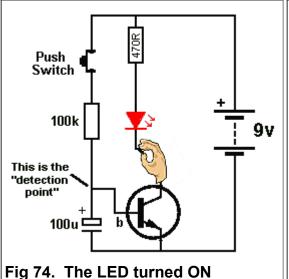
Here is an explanation of how the LED turns ON.
When the circuit is first assembled and the switch is not pressed, the transistor is not turned on and it is just like the diagram opposite. The LED is not connected to the transistor.
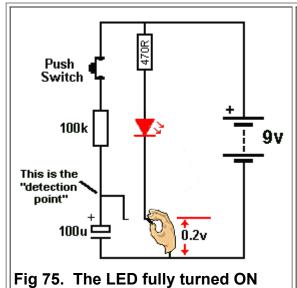
Fig 74. The LED turned ON

When the switch is pushed, the transistor turns **ON** (after a few seconds) and it pulls the lower lead of the LED down towards the 0v rail and this action turns the LED **ON**.

Fig 75. The LED fully turned ON

When the LED is **fully turned ON**, the lower lead of the LED is almost directly connected to the 0v rail.
In other words:
When the transistor is **FULLY TURNED ON**, the lower lead of the LED is almost directly connected to the 0v rail.
The voltage between the lead of the LED and 0v rail is 0.2v. This is the characteristic voltage across the collector-emitter terminals of a transistor when it is TURNED ON.

In the three diagrams above you can see the LED is changed from an **OFF** condition to an **ON** condition by the action of the transistor.
The transistor is acting **LIKE A SWITCH.**
This action is one of the most important actions in electronics.
It is called: **"The Transistor as a SWITCH"**
It is the basis to ALL Digital Circuits.
It is the basis because of these two facts:
1. When the transistor is **OFF**, the circuit is taking **no current** and no power is being lost or wasted.
2. When the transistor is **ON**, the LED is almost at 0v and no resistor is in the lower lead to waste any power.
Thus we can turn things **ON** and **OFF** without wasting and power.
This is the basis to **DIGITAL ELECTRONICS**.

DIGITAL ELECTRONICS revolves around circuits that are either **FULLY ON** or **FULLY OFF**. This means they take almost no power and we can combines lots of circuits and still take almost no power.
This means they do not get hot and it also means they will last a long time.
You may not think turning a transistor **ON** and **OFF** will achieve any worthwhile outcome but a circuit can be designed to use two transistors (similar to the **ROBOT MAN** above). The circuit does not Flip-Flop but requires a switch and when the switch is pressed, the circuit changes state. The two transistors are connected together and it takes two presses of the switch to make the output of the second transistor change state ONCE.
The circuit is a divider. It is called a: **divide-by-two** and is the basis of all counting in a computer.
By adding more "**divide-by-two**" circuits we can get "**divide by 4**, **divide by 8**" etc. Two transistors don't do much but when you combine millions of transistors we have a **COMPUTER.**

Two transistors can do one more thing. They can "**REMEMBER**."
Here is a manual circuit.
Pressing Switch **A** turns the LED **ON** and pressing switch **B** turns the LED **OFF**.
The circuit "remembers" or remains in each state called a stable state.

The technical name for this circuit is:
BISTABLE MULTIVIBRATOR or **BISTABLE SWITCH** or **BISTABLE LATCH**.

This is the basis to all the memory in a computer.
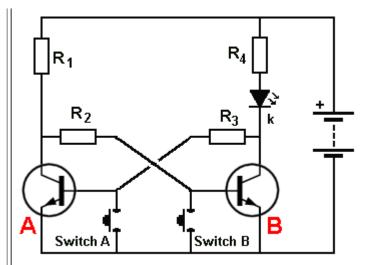
Fig 76. The "MEMORY CELL"

When Switch **A** is pressed, the voltage on the base is removed and transistor **A** turns OFF.
Transistor **B** turns ON via resistors R_1 and R_2 and the LED is turned **ON**.
When the switch is released, the voltage on the collector of transistor **B** is less than 0.6v and the two transistors remain in this state.
Pressing switch **B** turns the LED **OFF.** (transistor **A** turns ON via R_3, R_4 and the LED - very little current flows through the LED and you can hardly see it glowing). The voltage on the collector of transistor **A** is less than 0.6v and the two transistors remain in this state.

In electronics, we talk about the **DIGITAL TRANSISTOR** and **ANALOGUE TRANSISTOR**.
This is just an ordinary transistor (called a Bipolar Junction Transistor) in a **DIGITAL CIRCUIT** or **ANALOGUE CIRCUIT**.
We are now discussing the **DIGITAL CIRCUIT** - The Multivibrator - Astable Multivibrator and Bistable Multivibrator (Memory Circuit).
The **DIGITAL CIRCUIT** has **2 STATES**.
The **ON STATE** and the **OFF STATE**.
It is conducting in the ON STATE and the LED is illuminated.
In the OFF STATE, the LED is not illuminated.
In the ON STATE the transistor is said to be **CONDUCTING** or **BOTTOMED**.
In the OFF STATE the transistor is said to be "**CUT OFF** or "**OFF**."
These two states are reliable and guaranteed. They are not "half on" or "quarter on" or "75% off."
These states are easy to transmit "down a wire." The ON STATE is transmitted as "1" (voltage present) and the OFF STATE is transmitted as "0" (voltage not present).
These are the two **DIGITAL STATES**.

The **ROBOT MAN** is a **DIGITAL CIRCUIT**.
Each LED is ON or OFF.
The waveform on the output of each transistor is called a **DIGITAL SIGNAL**.
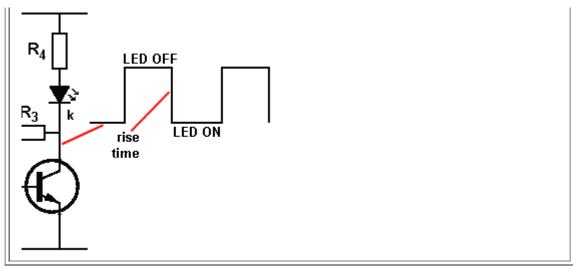
The waveform is said to be **DIGITAL** or **SQUARE WAVE**.
The top line of the graph represents the LED OFF.
The bottom line of the graph represents the LED ON. The LED is ON when the collector voltage is LOW because we are pulling the lead of the LED to the 0v rail as shown above.
The circuit changes from one state to the other very quickly and this is called the **RISE TIME**.
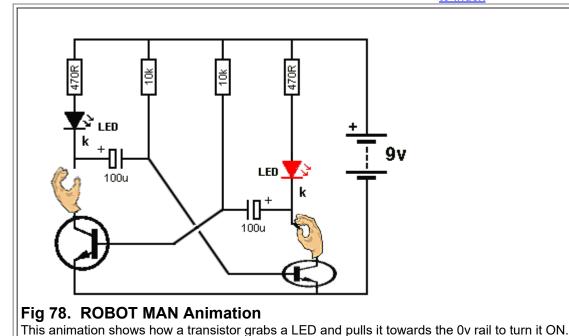
Going over the two **DIGITAL STATES** for a transistor.

In the first diagram the switch is not pressed and the base does not see a voltage to turn the transistor on. The transistor is "**OFF**" (not conducting) and it is not "grabbing" the LED. The LED is not illuminated.
In the second diagram the base of the transistor sees a voltage via the switch and it is TURNED on. The LED is illuminated.

Fig 77. The two Digital States

Fig 78. ROBOT MAN Animation
This animation shows how a transistor grabs a LED and pulls it towards the 0v rail to turn it ON.
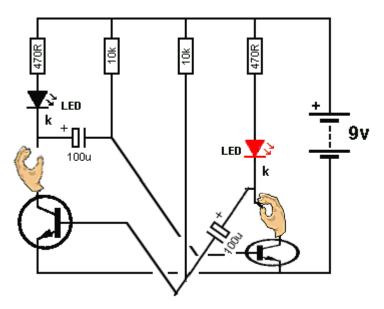
Fig 79. TIME DELAY Animation

The animation in **Fig 78** shows the two transistors turning the LEDs **ON** and **OFF** in a **FLIP FLOP** circuit.
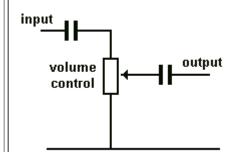
We know the 10k and 100u components form a **TIME DELAY** to create the time for each LED to be illuminated. The timing for one LED plus the other LED creates a **CYCLE** and this is the **FREQUENCY OF OPERATION** for the circuit. It is measured in cycles per second - Hertz - Hz.

We will now go into more detail of how the **TIMING COMPONENTS** create the **TIME DELAY** for each LED.

The circuit is more-complex than you think.

The 100u is already charged from a previous cycle and we show how it gets discharged via the 10k and charged in the opposite direction by the 10k to create a **TIME DELAY**.

THE CAPACITOR

The capacitor can perform many different functions and produce many different effects, depending on its value and the surrounding components.

In this circuit the capacitors on the input and output prevent DC on the volume control creating "scratchy sounds" when the volume is altered.

This is called "DC blocking."

The AC (the signal) passes through the capacitors but the DC voltage on the input is blocked.

CHARGING A CAPACITOR Part II
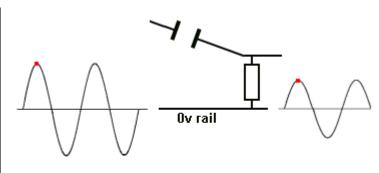
It is easy to see how a capacitor charges via a resistor in the **TIMING CIRCUIT** (Delay Circuit) above but many capacitors are not connected to the 0v rail.

They are connected as show in the animation below and their "job" is to pass a waveform.

When they pass the waveform they **CHARGE** and **DISCHARGE**.

The waveform is called an **AC SIGNAL** and the output is smaller than the input.

The circuit is taken from the circuit above, but the same effect applies to all capacitors that "pass a signal."

Here's why:
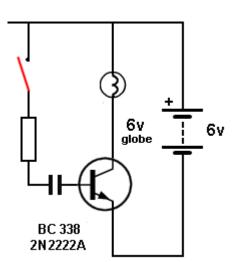
The capacitor charges slightly during the rise of the signal and the right-plate of the capacitor does not rise as high as the left-plate. That's why the output signal is not as large as the input signal.
If the capacitor did not charge, the output would be as large as the input. If you use a capacitor with a large value, it will not charge and thus the output will be as large as the input.
That's why you use a large capacitor !!!!

CHARGING A CAPACITOR Part III
Here is another CAPACITOR in action.

The animation shows a capacitor charging (via a resistor). The initial current is LARGE and this turns the transistor FULLY ON and the globe illuminates. As the capacitor charges, the base current reduces and the transistor starts to turn OFF. Eventually the capacitor is fully charged and the voltage on the base falls to 0v, turning the transistor OFF.
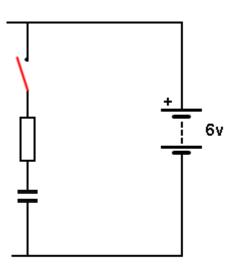
This animation shows three features:
1. The initial charging current is HIGH.
2. It gradually falls to zero.
3.The voltage on the base drops below 0.6v and the transistor turns OFF.
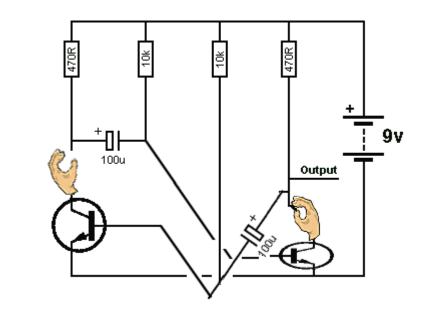
NEGATIVE VOLTAGE
You will be surprised to learn that many circuits produce a negative voltage or negative spike at some point (when doing circuit-analysis, each location or point or join of components is called a NODE) on the circuit. In other words the voltage will be LESS than the 0v rail of the circuit.
This is due to the presence of a capacitor and the animation shows how a capacitor can produce a negative voltage:

When a charged capacitor is "lowered from one position in a circuit" the positive lead may be lowered by say 3v. This means the other lead will be lowered by 3v. We are assuming the capacitor can be lowered and is not directly connected to the 0v rail.

You can see the electrolytic produces a NEGATIVE VOLTAGE on the base in the following animation, when the two transistors change states:
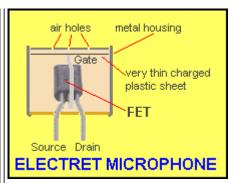
The Electret Microphone
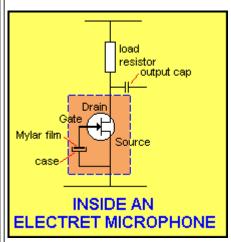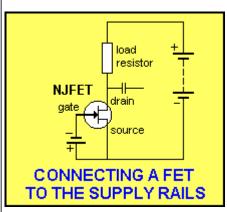
The most common type of microphone is the ELECTRET MICROPHONE.
It is incorrectly termed the "Capacitor Microphone" or Condenser Microphone." "Capacitor Microphone" descriptions make no mention of a FET as the amplifying device and a polarized diaphragm to detect the audio, so they are something different.
The electret microphone consists of a FET (transistor)
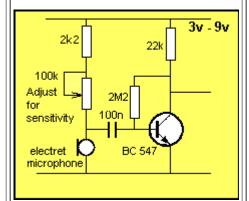
ELECTRET MICROPHONE

INSIDE AN ELECTRET MICROPHONE

CONNECTING A FET TO THE SUPPLY RAILS

inside an aluminium case with a very thin Mylar film at the front. This is charged and when it moves (due to the audio it receives via a small hole in the front of the case), it vibrates and sends a very small voltage to the GATE lead of the Field Effect Transistor. This transistor amplifies the signal and produces a waveform of about 2mV to 20mV at the output.

The electret microphone requires about 0.5mA and will operate from 1.5v supply with 4k7 LOAD RESISTOR. For 3v supply, the Load Resistor can be 22k to 47k. For higher supply voltages the resistor will be 68k or higher.

Electret microphones are extremely sensitive and will detect a pin-drop at 3 metres.

This lead connects to the case (negative)

Most electret microphones have two leads. One lead is connected to the case and this lead goes to the 0v rail. The other lead goes to a LOAD RESISTOR (4k7 to 68k - depending on the voltage of the project). Reducing the value of the load resistor will increase the sensitivity until the background noise is very noticeable.

They are used in Hearing Aids and are more-sensitive than the human ear.

They are very small, low-cost and very sensitive.

to Index

The Speaker

The most common speaker is about 30mm to 60mm diameter and 8 ohm impedance. This means the voice coil is about 8 ohms resistance.

The two leads can be connected either way to a circuit.

The speaker shown is 32mm diameter and has a realistic wattage of 100mW (NOT 1watt).

These speakers have a Mylar cone and the magnet is a "super magnet" and very small. That's why it is so flat.
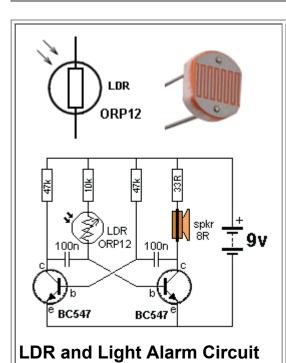
A speaker can be used as a microphone (called a Dynamic Microphone) and a circuit to connect the speaker (mic) to an amplifier can be found on Talking Electronics website. It is not as sensitive as an electret microphone and does not **Speaker Symbol** produce the same output amplitude, but it is an emergency microphone.

Light Dependent Resistor
(LDR)

Also called **PHOTOCELL** or **PHOTO RESISTOR**
A **Light Dependent Resistor** is a 2-leaded component containing a layer of semiconductor material.
The top contains two interleaving combs of conducting wires with a path of semiconductor material between. When light falls on the component, the resistance of the semiconductor material decreases.
In darkness the LDR will be about 300k. In very bright light the resistance will be about 200 ohms.
But if the light changes only a very small amount, the resistance CHANGE is VERY SMALL. For a large change, see **Photo Transistor**.
The **Light Alarm** circuit will produce a squeal when light falls on the LDR.
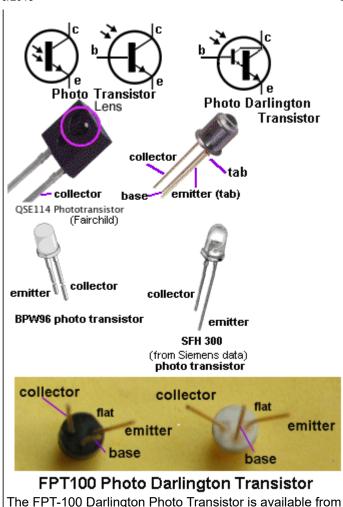
LDR and Light Alarm Circuit
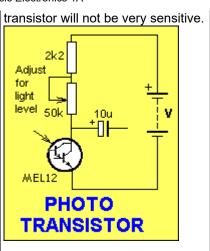
Photo Transistor

The **Photo Transistor** is very sensitive to changes in illumination. It is about 100 times more sensitive than the **LDR**.
The **Photo Transistor** is also available as a **DARLINGTON.**
The Darlington Photo-transistor is 100 x 100 times (10,000) more sensitive than the LDR.

The **Photo Transistor** and **Photo Darlington Transistor** are connected just like a normal transistor but the base lead is not connected. If the value of the LOAD RESISTOR is large, the

Photo Transistor
Photo Darlington Transistor

QSE114 Phototransistor (Fairchild)

BPW96 photo transistor

SFH 300 (from Siemens data) photo transistor

FPT100 Photo Darlington Transistor

The FPT-100 Darlington Photo Transistor is available from Talking Electronics for $1.00 each plus postage.

transistor will not be very sensitive.

2k2
Adjust for light level 50k
10u
MEL12
PHOTO TRANSISTOR

The Inductor

Also called **"coil,"** or **"Choke."**
An **Inductor** consists of one to many turns of wire wrapped around a former (tube of cardboard). The wire can be jumble-wound or wound in layers. The result is the same. This is called an **air-cored coil** or **air-cored inductor**.
The centre can be filled with a metal such as iron or laminations (thin sheets of metal) or a ferrite material. Different cores operate at higher frequencies.
The core can be circular (doughnut) or rectangular and it is called a **MAGNETIC CIRCUIT** (when it is a closed loop). Additional turns or increasing the diameter of the turns will increase the inductance.
A coil with a magnetic core can be used to pick up nails and metal items. It is called an **electromagnet**. It can be

Air-cored　　Air-cored　Iron-cored　Iron-cored
　　　　　　　variable　　　　　　　transformer

Ferrite-cored　　Ferrite-cored　Ferrite-cored
　　　　　　　　variable　　　　transformer

surface-mount
inductor 27uH
yes 27uH
270uH = 271

South
pole
North
pole

The emerging **magnetic lines of force** from an inductor produce the NORTH POLE - this is just "convention," a simple way to explain things - to get the explanation started. When two or more coils are wound near each other, the inductor is called a **TRANSFORMER** (but not the armature above).

operated on AC or DC.
When the metal core is loose and gets pulled into the coil it is called a **SOLENOID** or **ACTUATOR** or **LINEAR ACTUATOR**. It can be operated on AC or DC.
The way an inductor works is very complex but we can say it resists any rise or fall in voltage by turning the rise or fall into magnetic flux.
If the applied voltage is suddenly turned off, the inductor produces **a very high voltage** of **opposite polarity** (these are the **two** most important things for you to remember).

The Antenna
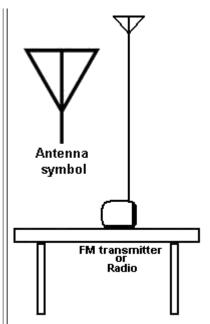
The first time you will need an antenna is when making an FM transmitter.
The antenna is usually a length of wire equal to half the wavelength of the transmitter. The frequency is about 100MHz and the wavelength is 3 metres.
A half-wave antenna is 1.5metres.
The length is important but the height is more important.
The wire should be as high as possible and "up-and-down" if the antenna on the radio is vertical, to get the maximum range.
This is called an **end-fed half-wave antenna** or **half-wave**
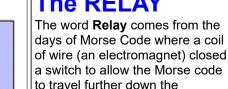
Monopole.
The transmitting circuit should have a good ground-plane such as connection to large batteries so the signal can be pushed and pulled into and out of the antenna.
This signal is then radiated as electromagnetic radiation to the surroundings.

Antenna symbol

FM transmitter or Radio

to Index

The RELAY
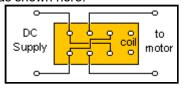
The word **Relay** comes from the days of Morse Code where a coil of wire (an electromagnet) closed a switch to allow the Morse code to travel further down the telegraph line.
It would "relay" or "pass-on" the information.
A relay allows a "weak circuit" (one with low current) to operate a LOAD that needs a large current. It also separates the two circuits electrically and prevents a voltage such as 240v connecting to a 12v circuit. The coil is separated from the contacts and this gives the two circuits isolation.
A double-pole double-throw relay can be used to reverse a motor as shown here:
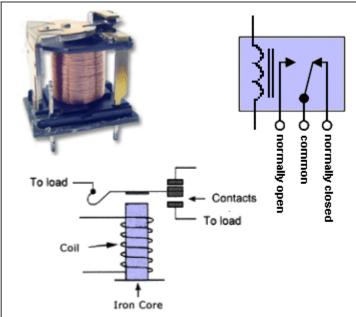
normally open   common   normally closed

To load

Contacts

To load

Coil

Iron Core

A single contact consists of 2 pins -called SPST (single-pole single-throw).
Or a single pair of contacts can consist of 3 pins - called change-over or SPDT (single-pole double-throw).
A double set of contacts consists of 6 pins, called DPDT (double-pole, double-throw). This is also called a **CHANGE-OVER RELAY** or **REVERSING RELAY** (when connected to a motor).
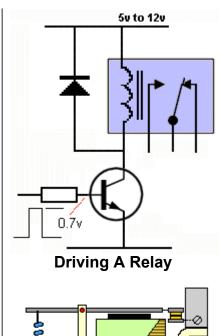
DC Supply   coil   to motor

to Index

Driving A Relay

(Powering A Relay)
The first thing you must decide is the voltage of the relay. This will depend on the voltage(s) available. The relay will be driven (activated) by a transistor and the base of the transistor only needs a signal (less than about 1v). This means the project can be operated on a voltage from 3v to 12v and the relay can be connected to a 5v to 12v supply.

Driving A Relay

The armature is drawn towards the coil when a current flows through the coil.
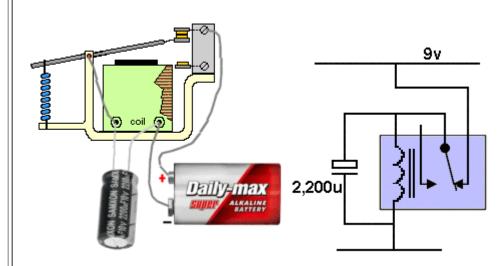
Next you need to know the current-rating of the contacts. This will depend on the current taken by the LOAD. The rating of most relays is: 1 amp, 5 amp or 10 amp.
Finally you need to know how many contacts are required.
For a single circuit you will need 2 pins and for two circuits you will need 4 pins (but relays only come with 6 pins).
You can get relays that need a very small current for activation. These are called CMOS relays. But most relays need about 100mA.
To protect the driving-transistor from spikes when the relay is turned off, you will need a diode across the coil.
The top animation shows a "single set of change-over contacts."
The lower animation shows the ARMATURE being drawn to the electromagnet. The electromagnet is the coil with a core of magnetic material that becomes a magnet (an electromagnet) when a current flows through the coil.

When the circuit is turned ON, the voltage across the 2,200u electrolytic is zero and it gradually charges. When the voltage is about 8v, the coil has enough voltage across it to pull the armature and open the contacts. The electrolytic supplies voltage to the coil for about 1 second and then the electromagnet does not have sufficient magnetism to hold the armature and it returns to close the contacts.

to Index

GATES

Animations of **GATES** and more details of their operation is covered in **DIGITAL ELECTRONICS** chapter.

Fig 80. The "AND" GATE and "OR" GATE with switches

The next **BUILDING BLOCK** we will cover is called the **GATE**.
In its simplest form it is an electrical circuit consisting of switches.
Its just two or more switches connected in series or parallel.
We give each circuit a name so we can talk about it and explain its action with a single word.
Later we will cover the electronic version and show how diodes and a transistor are needed to perform a **GATING FUNCTION.**
The type of GATE we are talking about is a **LOGIC GATE**.
The circuit performs an operation called a **LOGICAL OPERATION** on an input or a number of inputs and creates a single output - called a **LOGICAL OUTPUT**.
LOGICAL means "understandable" or "correct" and in this case it means DIGITAL - the signal will rise to full rail voltage or fall to zero voltage. The output will not be half rail or quarter-rail voltage.
The diagram show an **"AND" GATE** and **"OR" GATE** with switches.
For the **AND GATE** close switch A **AND** switch B for the lamp to illuminate.
For the **OR GATE** close switch A **OR** switch B for the lamp to illuminate.

Fig 81. The "NAND" GATE and "NOR" GATE with switches and a transistor

INVERSION
Inversion produces the opposite effect to the results above.
Suppose we want to turn OFF a lamp when one or two switches are pressed. We need a transistor.
The technical word for Inversion is **NOT**. It is simplified to the letter "**N**."
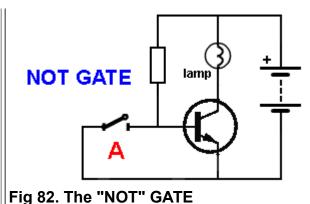
For the **NAND GATE** close switch A **PLUS** switch B for the lamp to turn OFF.
For the **NOR GATE** close switch A **OR** switch B for the lamp to turn OFF.

These gates are only **demonstration-gates** to show how one or two switches will turn a lamp OFF.

NOT GATE

A single switch and transistor produces a **NOT GATE.**
This is simply an **INVERSION**.

The resistor turns the transistor **ON** and the lamp illuminates. The switch removes the voltage on the base and the transistor turns **OFF**.
This is only a demonstration circuit to show how a switch can turn a lamp OFF.

Fig 82. The "NOT" GATE

The 5 gates above form the basis to turning a circuit **ON** and **OFF**. We will discuss these gates later in the digital section.

The next building block is the **GATING DIODE**.
We have shown a diode allows current to flow when the diode is correctly placed in a circuit and blocks current when it is reversed.
The 5 gates above are electrical circuits but an electronic circuit works in a slightly different way.
The electronic circuit will be covered later in the DIGITAL section.
For the moment we will explain how a diode can be used to create a **GATE**.
In other words it creates a ONE-WAY PATH to allows signals to pass from one stage to another and prevents signals passing in the opposite direction.
In the AND GATE circuit, both inputs are LOW and current flows through the resistor (it will get HOT). When one input is taken HIGH, current still flows through the other diode and the lamp does not illuminate.
When BOTH inputs are HIGH, current flows through the resistor to illuminate the lamp. No current flows through the diodes.
In the OR GATE, when one input is taken HIGH, current flows through the diode to illuminate the lamp. This can be done with EITHER input.
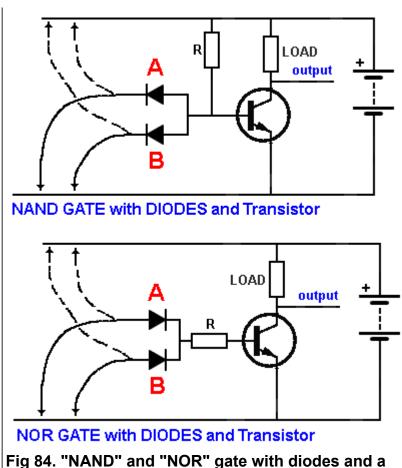
AND GATE with DIODES

OR GATE with DIODES

Fig 83. "AND and "OR" gate with diodes
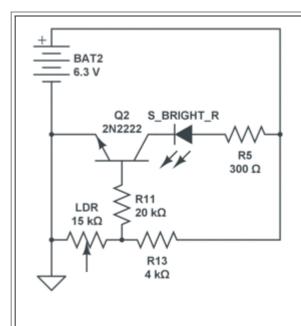
A **NAND** and **NOR** gate can be made with diodes and a transistor. This time we the output is either HIGH or LOW. We are gradually producing circuits that

NAND GATE with DIODES and Transistor

NOR GATE with DIODES and Transistor

Fig 84. "NAND" and "NOR" gate with diodes and a transistor.

are electronic, rather than electrical circuits.
In the **NAND GATE** circuit, taking one of the inputs HIGH will still allow the other input to prevent the transistor turning ON.
When BOTH inputs are HIGH, the transistor turns on via resistor R and the output is LOW.

In the **NOR GATE** circuit, taking one of the inputs HIGH will turn the transistor ON and the output will be LOW.
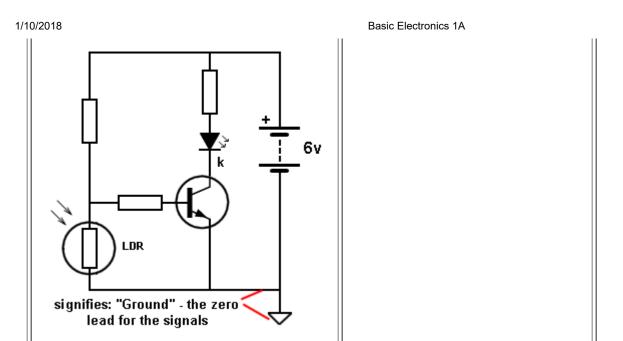
Drawing A Circuit

A circuit must be drawn according to simple rules so it can be instantly recognised.
An electronics engineer can "**see a circuit working**" when it is drawn correctly and can see if it is drawn correctly; if the parts-values are correct and can use the circuit to assist in diagnosing a problem with a faulty circuit.

The top circuit on is very difficult to visualise because it is not drawn in the normal way.
All the components have to be "**turned around in your mind**," to see what the circuit is doing.

signifies: "Ground" - the zero
lead for the signals

QUICK QUIZ - to see how much you know

following 50 questions . . . JavaScript is required!
This test will see how much you have learnt.

An open circuit   help

**4. What is the approximate characteristic voltage that develops
across a red LED?**

☐ 1.7v
☐ 3.4v
☐ 0.6v
☐ 5v   **help**

5. If two resistors are placed in series, is the final resistance:

- ☐ Higher
- ☐ Lower
- ☐ The same
- ☐ Cannot be determined   **help**

6. Which is not a "common" value of resistance:

- ☐ 2k7
- ☐ 1M8
- ☐ 330R
- ☐ 4k4   **help**

7. Which value of resistance, placed across a 9v battery
   will get hot:

- ☐ 22k
- ☐ 22R
- ☐ 220k   **help**

8. If the voltage on the base of a transistor increases, does it:
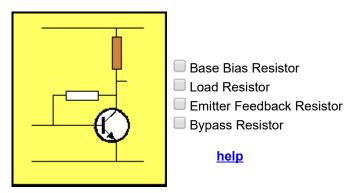
- ☐ Turn on
- ☐ Turn off
- ☐ Not enough information
- ☐ Remain the same   **help**

9. The resistor identified in brown is called the:

- ☐ Base Bias Resistor
- ☐ Load Resistor
- ☐ Emitter Feedback Resistor
- ☐ Bypass Resistor

      help

10. The first three colour bands on a resistor are:
   yellow - purple - orange

- ☐ 47k
- ☐ 4k7
- ☐ 470k
- ☐ 4R7   **help**

11. A resistor with colour bands: red-red-red-gold, has the value:
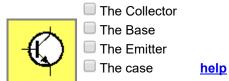
- ☐ 22k 5%
- ☐ 2k2 5%
- ☐ 220R 5%
- ☐ 22R 5%   **help**
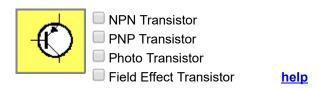
12. The lead marked with the arrow is:

☐ The Collector
☐ The Base
☐ The Emitter
☐ The case   **help**

13. A 10k resistor in parallel with 10k produces:

☐ 10k
☐ 5k
☐ 20k
☐ Cannot be determined   **help**

14. The symbol is:

☐ NPN Transistor
☐ PNP Transistor
☐ Photo Transistor
☐ Field Effect Transistor   **help**
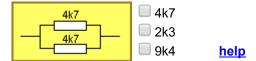
15. Two 3v batteries are connected as shown. The output voltage is:

☐ 3v
☐ 0v
☐ 6v   **help**
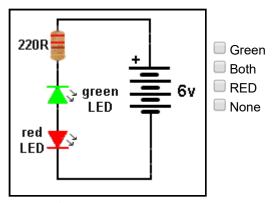
16. 4 resistors in ascending order are:

☐ 22R   270k   2k2   1M
☐ 4k7   10k   47R   330k
☐ 3R3   4R7   22R   5k6
☐ 100R   10k   1M   3k3   **help**
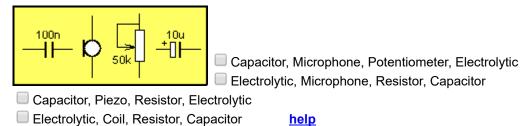
17. The closest value for this combination is:

☐ 4k7
☐ 2k3
☐ 9k4   **help**
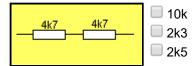
18. Which LED will illuminate:

☐ Green
☐ Both
☐ RED
☐ None

19. The four symbols are:

☐ Capacitor, Microphone, Potentiometer, Electrolytic
☐ Electrolytic, Microphone, Resistor, Capacitor

☐ Capacitor, Piezo, Resistor, Electrolytic
☐ Electrolytic, Coil, Resistor, Capacitor   **help**
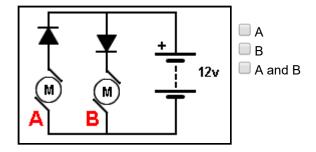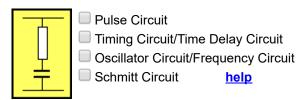
20. The closest value of the combination is:

☐ 10k
☐ 2k3
☐ 2k5

21. Which motor will work:

☐ A
☐ B
☐ A and B

22. A resistor and capacitor in series is called a:

☐ Pulse Circuit
☐ Timing Circuit/Time Delay Circuit
☐ Oscillator Circuit/Frequency Circuit
☐ Schmitt Circuit   **help**

23. A red-red-red-gold resistor in series with an
orange-orange-orange-gold resistor produces:

☐ 5k5
☐ 35,200 ohms
☐ 55k
☐ None of the above   **help**

24. Name the 4 components:
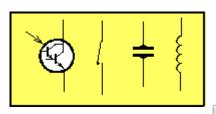
☐ Photo transistor, switch, capacitor, coil
☐ Transistor, mercury switch, piezo, inductor
☐ Photo transistor, reed switch, piezo, coil
☐ Photo Darlington transistor, switch, piezo, inductor   **help**
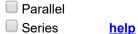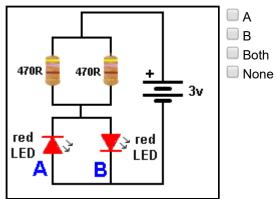
25. To obtain a higher value of resistance, resistors are
connected in:

☐ Reverse
☐ Forward

☐ Parallel

☐ Series   **help**
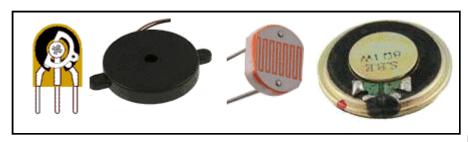
26. Which LED will illuminate:

☐ A

☐ B

☐ Both

☐ None

27. Name the component that detects light:

☐ mini trim pot

☐ Light Dependent Resistor

☐ piezo

☐ speaker   **help**

28. What is 1,000p?

☐ 0.01n

☐ 0.0001u

☐ 0.1n

☐ 1n   **help**

29. The current in a circuit is 45mA. This is:

☐ 0.045Amp

☐ 0.00045A

☐ 0.0045A

☐ 0.45A   **help**

30. A 100n capacitor can be expressed as:

☐ 0.1u   u = microfarad

☐ 0.01u

☐ 0.001u

☐ none of the above   **help**

31. 1mA is equal to:

☐ 0.001A

☐ 0.00001A

☐ 0.01A

☐ 0.1A   **help**

32. 1,200mV is equal to:

☐ 12v
☐ 1.2v
☐ 0.12v
☐ 0.0012v   **help**
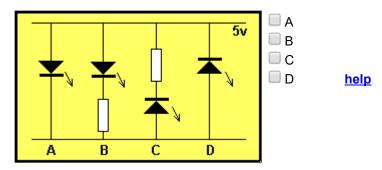
33. The approximate current for a toy 3v motor is:

☐ 10mA
☐ 100mA to 300mA
☐ 1 amp
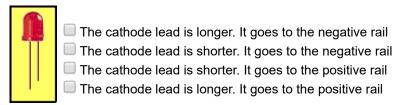
34. What is the resistance of this resistor:

☐ 47k
☐ 4k7
☐ 4R7

35. Identify the correctly connected LED:

☐ A
☐ B
☐ C
☐ D   **help**

36. Identify the correct statement:
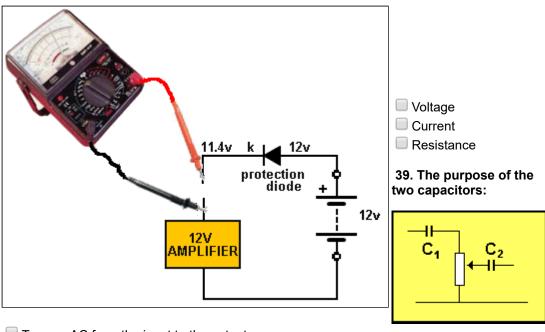
☐ The cathode lead is longer. It goes to the negative rail
☐ The cathode lead is shorter. It goes to the negative rail
☐ The cathode lead is shorter. It goes to the positive rail
☐ The cathode lead is longer. It goes to the positive rail   **help**

37. The current requirement of a LED is:
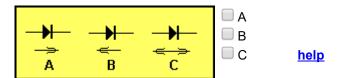
☐ 1.7mA
☐ 25mA
☐ Between 3 and 35mA
☐ 65mA   **help**

38. The multimeter is measuring . . .

☐ Voltage
☐ Current
☐ Resistance

39. The purpose of the two capacitors:

☐ To pass AC from the input to the output
☐ To allow the signal to oscillate
☐ To pass DC from the input to the output
☐ To amplify the signal   **help**

40. The direction of conduction for a diode is:

☐ A
☐ B
☐ C   **help**

41. A DC voltage . . .

☐ rises and falls
☐ is a sinewave
☐ remains constant
☐ is an audio waveform   **help**

**42. Arrange these in ascending order: k, R, M
(as applied to resistor values)**

☐ R, k, M
☐ M, R, k
☐ k, M, R
☐ M, k, R   **help**

43. A battery produces AC current:
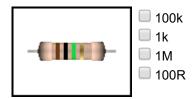
☐ true
☐ false

44. The tolerance bands: gold, silver, represent:
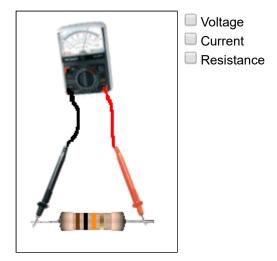
☐ 5%, 10%
☐ 10%, 5%   **help**

45. 223 on a capacitor represents:
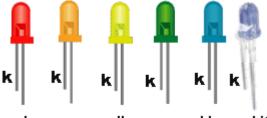
☐ 0.022u   u = microfarad
☐ 22n   n = nanofarad
☐ 22,000p   p = picofarad
☐ All of the above   **help**

46. Arrange these in ascending order: n, p, u
    (as applied to capacitor values)

☐ p, u, n,
☐ n, u, p
☐ p, n, u   **help**

47. What is the resistance of this resistor:

☐ 100k
☐ 1k
☐ 1M
☐ 100R

48. The number "104" on a capacitor indicates:

☐ 0.1u
☐ 100n
☐ 1n
☐ 10n   **help**

49. What is the multimeter detecting:

☐ Voltage
☐ Current
☐ Resistance

50. For the LEDs, what is the characteristic voltage for the red and white LEDs:

red   orange   yellow   green·   blue   white

☐ 3.6v, 1.7v
☐ 2.4v, 3.3v
☐ 1.7v, 3.6v

☐ Cannot be determined

Final Assessment

THE CRYSTAL-SET RADIO

100 more Crystal Set plans

Everyone wants to make a RADIO.

The simplest radio is a **CRYSTAL SET**. (Sometimes called a **Crystal-Set Radio** or **Xtal Radio Set** or **Crystal Diode Radio**.)

However a **Crystal Set** needs a number of components that are very hard to get: (tuning capacitor with knob) and (crystal earpiece for $1.25).
The "air" TUNING CAPACITOR (20p to 415p) is not easily available and the postage is expensive. The germanium diode is a special component and the aerial coil wound on a ferrite slab is difficult to obtain.

But you don't need these components. They can be substituted.

There are hundreds of websites on the internet describing the **CRYSTAL SET** and if you want to build a "normal" set, you can Google these sites or **100 more Crystal Set plans**. Many of them sell kits too.

But this article is different.

We are going to have all the fun of making a **CRYSTAL SET** but with modern components and easy-to-make components and with an amplifier stage. The output is loud so you don't need a long antenna. And we are going to make our own TUNING CAPACITOR and a very simple aerial coil (called a FRAME AERIAL) as well as replacements for the germanium diode (use a TRF radio IC or a transistor) and in place of hi-impedance headphones (use a piezo diaphragm) and a crystal earpiece equivalent (a piezo diaphragm).
It's even better to have one of each type of component so you can compare the performance, so no matter how many parts your get, nothing will be wasted.

 We are also going to explain the fundaments of how the circuit works as even the simplest circuit has a number of very important features that are used in many other circuits.

But first we are going to learn about the components and how they combine to make the circuit work.
When two or more components are connected together they sometimes produce a completely different result to the capabilities of either item.
This is the case with a capacitor and inductor in parallel. An inductor is simply a coil - turns of wire on a cardboard tube - called a former and the centre of the coil is AIR. It is called an air-cored coil or air-cored inductor.
Each component (the coil and capacitor) is called a PASSIVE DEVICE - in other words it does not amplify, but when the are connected together they create a result very near to amplification. And they also produce a result of picking up a huge number of signals and only allowing one signal to appear across the pair. A truly amazing result.

We start by placing a capacitor across the coil.
There is so much activity in the air, from radio, TV, taxi and mobile phone usage that the air is filled with electromagnetic radiation.
This radiation will cut the turns of the inductor (the coil) and produce a microscopic voltage in the turns. This is enough to start the two components passing energy back and forth at a rate

determined by their values. This is the basis of our first discussion.
These two components are called a TUNED CIRCUIT and make up our first building block called THE FRONT END.
.

THE FRONT END

This consists of a coil and capacitor. These two components are in PARALLEL and the signal (called the RADIO SIGNAL or RADIO WAVE) passes through the centre of the coil and produces a voltage in the turns of the coil. The electromagnetic wave has to pass through the centre of the coil.
This voltage is the result of a mass of signals that are interfering with each other and producing a signal called BACKGROUND NOISE.
The voltage can be increased by an external aerial (called an ANTENNA) and it consists of ALL THE LOCAL radio stations (and everything else).
The voltage is a mass of signals and is absolutely useless as it represents all the stations AT THE SAME TIME.
However, across the coil is a capacitor and these signals charge the capacitor with the very small voltage produced by the energy of the signals. When the capacitor is charged, it delivers its voltage to the coil. The coil accepts the energy and converts it to magnetic flux.
After a very short period of time the capacitor becomes discharged and the magnetic flux collapses and produces a voltage in the coil of the opposite polarity to charge the capacitor again in the opposite direction.
These two components keep oscillating back and forth, using the tiny amount of energy from the stations.
There is a natural frequency for the capacitor and coil to pass energy back and forth and one of the stations will provide energy to assist this natural frequency.
When this happens, the amplitude of the signal increases and the signals from all the other stations cancel themselves out and only one signal (waveform) remains.
This signal is called the NATURAL FREQUENCY OF RESONANCE and it corresponds exactly to one of the radio stations.

The end result is a waveform that is the exact same frequency as one of the radio stations and when voice or music is played, the amplitude of the waveform increases and decreases. This will be the signal you hear in the earpiece or speaker.
It is called an AMPLITUDE MODULATED signal and that is where we get **AM RADIO** from.

Understanding the concept of a CAPACITOR and INDUCTOR in parallel is very important. They form a TUNED CIRCUIT that has a natural RESONANT FREQUENCY.
Here is a very similar analogy. You have a heavy metal ball on a long string attached to the gutter on your house - just like a pendulum. You can push the ball very lightly with a finger and after a number of pushes you will be able to get the heavy ball swinging in a very large arc.
The only way to keep it swinging is to push it very lightly at exactly the right time. If you push it at the wrong time it will eventually stop swinging.
The parallel tuned circuit is exactly like the ball. It wants to oscillate at a particular frequency.
All the radio stations are pushing and pulling the circuit at the wrong times and nothing is happening. But one radio station pushes at exactly the right time and the circuit starts to oscillate.
All the other stations are fighting each other just like one person pushing the ball sideways and another pushing the ball from the opposite side. The results cancel each other and you are the only one assisting the swing.
The TUNED CIRCUIT can also be called a FILTER with a very narrow BAND-PASS frequency but our simple explanation describes the operation much more clearly.

There are a few other terms used to describe the components in the font end:

LOOP STICK ANTENNA -This is an alternate name given to the coil of wire wound on a ferrite rod or slab. It also has the name ROD ANTENNA or FERRITE ROD ANTENNA.
The winding can be enamelled wire or flexible wire called LITZ WIRE. This is very fine strands of enamelled wire twisted together and covered in cotton. The purpose of changing a thick wire to lots of very thin wires is to prevent the radio signals creating loops of signals within the wire and these signals will cancel each other and not produce a signal out the end of the wire.
In our experiments, we have not noticed any difference in a coil made with ordinary enamelled wire and Litz wire.
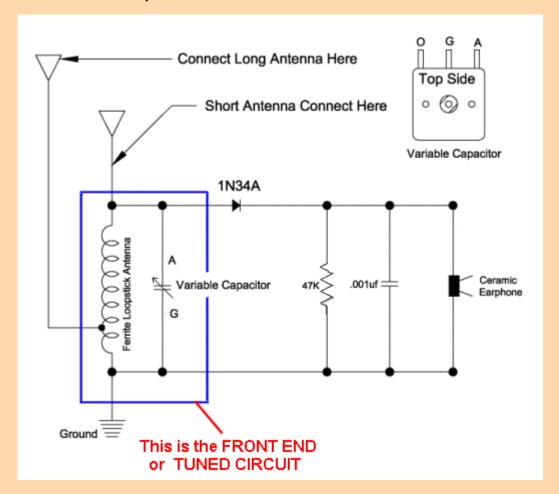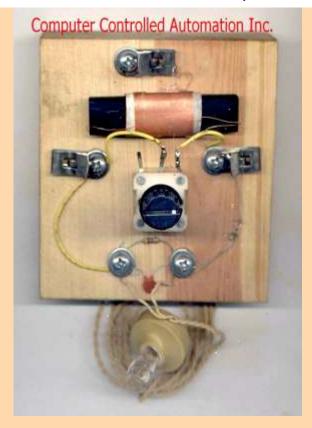
Here is a set of components to make your own Crystal set from **Scott's Electronic Parts**:
You can see the rod antenna, germanium diode, crystal earpiece, capacitor and resistor. The kit costs about $9.00 plus postage and includes knob, clips and screws but no board to mount the parts.

Here is the circuit for a Crystal Set:

Connect Long Antenna Here

Short Antenna Connect Here

O   G   A

Top Side

Variable Capacitor

1N34A

Ferrite Loopstick Antenna

A

Variable Capacitor

G

47K

.001uf

Ceramic Earphone

Ground

This is the FRONT END
or TUNED CIRCUIT
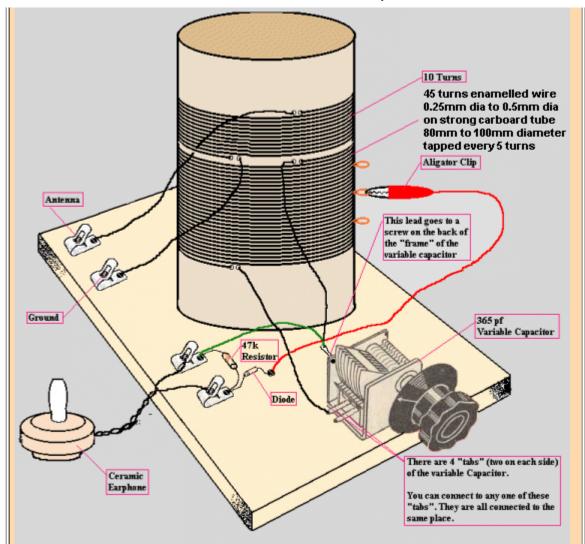
Here are the components mounted on a board called BREADBOARD:

The top clip connects to a long antenna. The left clip connects to ground and the right clip connects to a short antenna.

Let me clear up a point. You do not need a ferrite rod antenna for the coil. You can use an ordinary coil wound on a cardboard tube and it will work just as well if you are using an outside antenna.
Here is the circuit using a home-made coil.

10 Turns

45 turns enamelled wire
0.25mm dia to 0.5mm dia
on strong carboard tube
80mm to 100mm diameter
tapped every 5 turns

Aligator Clip

Antenna

This lead goes to a
screw on the back of
the "frame" of the
variable capacitor

Ground

365 pf
Variable Capacitor

47k
Resistor

Diode

There are 4 "tabs" (two on each side)
of the variable Capacitor.

You can connect to any one of these
"tabs". They are all connected to the
same place.

Ceramic
Earphone

The tappings on the coil allow a wide band of radio stations to be tuned.
Each tapping allows a different portion of the band to be covered.

The next part to understand is this:
The coil and capacitor must not be LOADED. In other words, you cannot connect anything to this combination because the signal it is producing will be "taken away" or "removed" or "considerably reduced" by the item you are connecting to the circuit.
These two components are called a TUNED CIRCUIT and when they are not loaded they pick up all the radio stations, one station at a time, when the natural resonant frequency of the coil and capacitor exactly match the frequency of the radio station. The circuit actually "rejects" all the radio stations except one. Because all the other stations are trying to make the Tuned Circuit oscillate at a different frequency and it does not do this.
The result of the TUNED CIRCUIT oscillating under NO LOAD conditions produces a waveform that is very high and this gives the circuit GOOD SELECTIVITY. The circuit can select one station and reject nearby stations.
It also has good SENSITIVITY as it can pick up weak stations.
If you load the circuit, only the strongest signal will be detected and it will be spread across the full range of the tuning capacitor.
Obviously the theory is more-complex but we are explaining the end-result.
Theory talks about the "Q" value of the coil and this is its ability to produce a very good output when the magnetic flux collapses and the "Q" value increases when the circuit is not loaded. .
Although these voltages are very small (in the order of microvolts or millivolts) the result is very important as the rest of the circuit will be amplifying this waveform a few thousand times.
As we explained above, pushing the weight on a string only needs a push of 1 cm and eventually the weight will swing 1 metre. This is a gain of 100:1 The same thing happens with the tuned circuit. The incoming radio signal is in the order of microvolts, but the coil and capacitor will produce a signal as high as 500 millivolts. This is an improvement or "gain" of more than 1,000 and is referred to as the "Q" of the circuit.
You will also notice the TUNED CIRCUIT is not connected to any supply voltage. It does not have

be connected. It generates its own waveform from the signals in the air.
It should not have any DC current flowing through it via the supply as this would put a load on the circuit and reduce its operation.

However we must "pick-off" the signal so it can be amplified.
This must be done with a very high impedance circuit.

THE CONVERTER (detector) - THE DIODE

The next part of the circuit is the CONVERTER. Commonly called the DETECTOR.
It converts the RADIO FREQUENCY signal to an AUDIO FREQUENCY signal.
This is the job of the DIODE.
The radio frequency signal is a very high frequency signal (say one million cycles per second) and it is sending a tone of one thousand cycles per second through the air-waves.
What is happening is this: The one megahertz signal has a certain amplitude and over a range of the first one-thousand cycles, the amplitude gradually decreases and then increases again. If you look at the tops of this 1,000 cycles you will see a waveform that corresponds to the one kilo-Hertz signal.
The 1MHz signal is picked up by the coil and capacitor in the front end and makes it oscillate.
The radio frequency signal is gradually getting larger over 500 cycles then smaller over the next 500 cycles and this increase and decrease represents the 1,000 cycles per second tone.
This is the waveform (the signal) that passes through the diode.   This will be explained further in a moment.
The diode does not pass any signals less than 200mV as the first 200mV is lost in the junction of the diode. This means the signals start to appear on the other end of the diode when they are above 200mV.

This is how the diode works:
Across the crystal earpiece is a capacitor. The capacitor gets charged via the diode.
The diode is present to stop the capacitor getting discharged when the waveform is in the wrong direction. (by this we mean - when the waveform is lower or smaller in amplitude than the voltage on the capacitor).
And the waveform is in the wrong direction about 50% of the time. To charge the capacitor for one-half-cycle requires 500 "little increments" in voltage with each increment adding a microscopic increase in voltage. We don't want this voltage to reduce when the waveform is reversing direction and the diode stops the voltage flowing back to the Tuned Circuit.
During the next half of the cycle when the pulses are getting smaller and smaller, the voltage on the capacitor is "bled off" by the load resistor.
The crystal earpiece detects this voltage. What we mean, is the diode allows the voltage to rise (increase) on the capacitor via lots of little "pulses" and the voltage increases in the form of a sinewave to a maximum amount. This voltage is passed to the crystal earpiece.
Once the voltage rises to a maximum, the little pulses of energy are not quite as strong, and the voltage on the capacitor reduces to form the second portion of the sinewave. This voltage is always being passed to the crystal earpiece and you can hear it as an audio signal.

GERMANIUM OR SILICON DIODE

The preferred type of diode for a Crystal Set is germanium. This is because it drops only about 0.3v.
But a silicon diode can be used, even though it drops about 0.7v, if the radio stations are very loud (close by).
You have to remember, you need a very good aerial (and a water-pipe earth) to get any results with a Crystal Set because you are asking the signal to provide the energy to drive the earpiece.
By simply adding a transistor, you are improving the performance 100 times and the long antenna can be reduced to a FRAME ANTENNA and the earth can be the metal frame of your soldering iron.
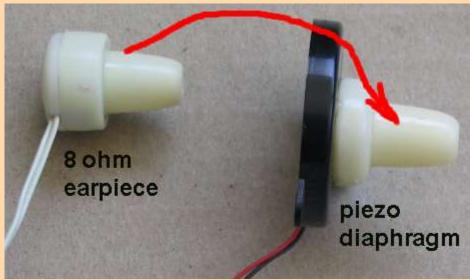
THE EARPIECE or EARPHONE
also The Magnetic Earpiece or CRYSTAL EARPIECE

The earphone or earpiece used in a Crystal Set must be a high impedance device because the crystal set does not produce a high current and cannot drive a low-impedance earpiece. That's why a CRYSTAL EARPIECE is ideal.
It has a crystal glued to the back of the earpiece and connected to its top surface is an aluminium diaphragm. When the crystal expands and contracts as a result of a voltage applied via two electrodes, the diaphragm moves and you can hear the signal. It exhibits a very high impedance

because it consists of a crystal and no coil of wire is contained inside the case.
If you do not have a Crystal Earpiece, you can make your own from the shell of an 8 ohm earpiece and a piezo diaphragm. Only the front part of the earpiece is used.

Make your own Crystal Earpiece

Hit the 8 ohm earpiece on the side and the front comes off. Glue the front onto a piezo diaphragm with hot-melt glue. See photo above.
The piezo diaphragm is a ceramic substrate that deflects in the presence of a voltage. It is quite sensitive and you can hear the audio quite clearly.
The waveform emerging from the diode in a Crystal Set is called AUDIO and although it has an amplitude of a few hundred millivolts, it does not have any current associated with it. The crystal earpiece and the piezo diaphragm react to this voltage.

THE 8ohm EARPIECE
The 8 ohm earpiece can be used with our 8ohm Buffer stage shown below.

16ohm 32 ohm and 64 ohm EARPIECE(s)
Earpieces and headsets from mobile phones are 16 ohm or 32 ohm per earpiece and are terminated via a stereo 2.5mm or 3.5mm plug. The earpieces are connected in SERIES to get the best coupling to our radio circuits and you need to find the two pins on a stereo socket to produce series connection. Get a multimeter and switch to "ohms." Try all the pins and you will get a click in the left ear then the right ear. Keep searching until you get a click in both earpieces at the same time. Use these two pins.

Stereo mobile phone headset - unusually 32R or 64R

PROBLEMS
The biggest problem with a Crystal Set is the need for a long antenna.
The first 200mV to 300mV of a signal is lost in the diode and you need a long antenna to pick up a signal so the output of the TUNED CIRCUIT has enough voltage to drive the high-impedance earpiece.
This requires an outside aerial 5 metres long and 3 metres high.
This is not practical for most hobbyists so we will be adding an amplifying stage to the crystal set so a shorter (smaller) aerial can be used.
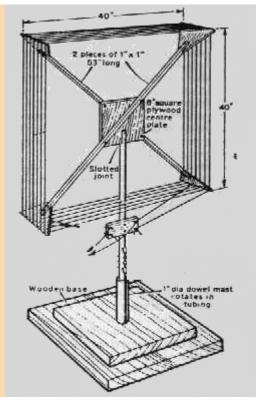
THE FRAME AERIAL or FRAME ANTENNA or FRAME COIL
The aerial coil shown in the photo above is a ferrite slab with about 80 turns of Litz wire. You can find one of these in an old broken AM radio or from a parts-shop. In the instructions below we show how to make your own Ferrite Rod Antenna
An equally-good substitute is a frame antenna made by winding insulated wire in a rectangle around wooden sticks.
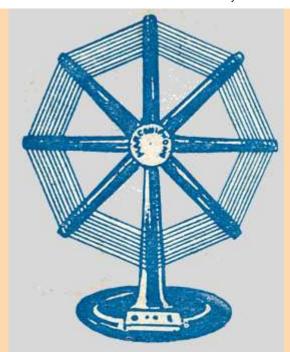
FRAME ANTENNA

15 turns on a diamond frame

One of the earliest Frame Antennas

The Frame Aerial can be as large as 100cm x 100cm or as small as 10cm x 10cm around a plastic chocolate box.

Here are two FRAME ANTENNAS:

This will work just as good as a ferrite slab antenna. The slab antenna is just 100 times smaller. The slab antenna was invented so a transistor radio could be built in a small case. But if it is not available, you can wind 20 turns around a plastic chocolate box and it will work just as good.

Alternately you can wind 20 turns around a biscuit tin. Put a pencil on the tin and wind the turns over the pencil too. Remove the pencil and it will be easy to remove the turns. Use tape to keep

the turns together.

The FRAME AERIAL does two things. It picks up the radio waves and it becomes the coil (called the INDUCTOR) in the TUNED CIRCUIT. It must be placed away from metal objects, such as a refrigerator.

BASKET WEAVE COIL

There is no point making a complex BASKET WEAVE COIL as it will not work any better than simply jumble winding all the turns at the maximum circumference of the coil, because the energy capturing capability of the coil relies entirely on the amount of flux lines passing through the centre of the coil.
By increasing the centre of the coil, the amount of flux is increased for the same coil size.
In fact, the simplest and cheapest is to wind turns around a box, as explained later in this article, or make a frame antenna as shown above. Technically speaking, a round coil has the best performance but only by a few percent.

A BASKET WEAVE COIL

THE VARIABLE INDUCTANCE TUNING COIL

Whenever the size or shape of the coil is changed, (or the number of turns), the natural frequency of the Tuned Circuit will change and a different radio station will be picked up.
This means tuning across the band can be done by altering the characteristics of the coil while keeping the value of the capacitor fixed.
Changing the inductance can be done in many different ways.
The coil can have taps every 5 turns and an alligator clips selects the correct tap. But very few radio stations will correspond exactly to each tap.
Another way is to have a slider move up and down the turns as shown in the following image:

The slider makes contact where the insulation has been removed. But it may touch two turns at the same time and create a "shorted turn" and reduce the "Q" of the coil.

Another way is to move a ferrite bar (rod) in and out of the coil:
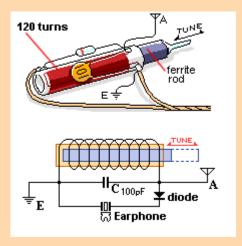
THE SLUG TUNED COIL

To tune across the radio band, the natural frequency of oscillation of the TUNED CIRCUIT must be adjusted (changed). This can be done by changing the value of the capacitor or the value of the inductor.

The value of the inductor can be changed by adding or removing turns or changing the amount of magnetic material in the centre of the coil.
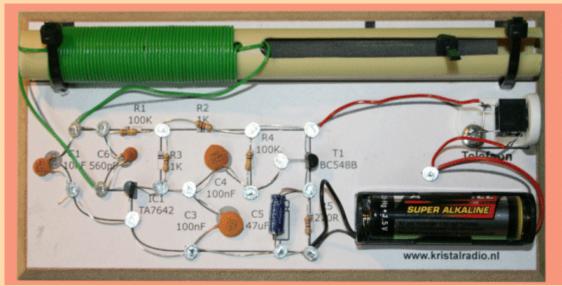
A ferrite bar can be screwed in and out of the coil or slid in and out and this component is called a **SLUG TUNED COIL**.

The following diagram shows a **SLUG TUNED CRYSTAL SET**:
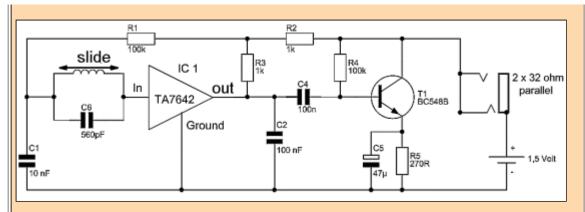
By changing the value of the 100p capacitor, different parts of the band can be picked up.

The photo shows a slug tuned coil using 60 turns of insulated wire on a 10mm tube (or any tube that will fit over a 8-10mm ferrite rod) and a circuit containing an AM radio chip plus a buffer driver transistor:

A SLUG-TUNED RADIO

The circuit above is has a broad-band amplifier consisting of 10 transistors (IC1) and they are directly coupled (connected) to each other because it is not possible to "manufacture" a capacitor inside the IC. The IC has 3 terminals (pins, legs) and it looks like an ordinary transistor. Experimenting with this type of IC has shown that it is no better than 2 ordinary transistors connected in a direct-coupling arrangement.

Here is the address of the site for the slug-tuned radio.
http://www.kristalradio.nl/
Unfortunately the site is in Dutch and the kit is not available. However the photos give a clear picture of the how the parts are connected.

The inductance of the coil can also be altered by winding another coil and placing it near the first coil so that the magnetic field interacts with each other and changes the inductance of the circuit. This is called a VARIABLE INDUCTANCE TUNING COIL.
You can have one coil inside the other, two coils near each other or two flat coils side-by-side. Any two coils will interact with each other.
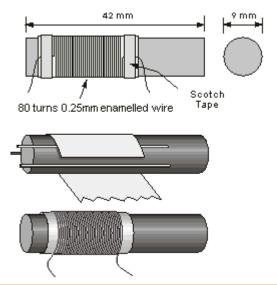
An Inductive TUNING COIL called a VARIOMETER

MAKING YOUR OWN FERRITE ROD ANTENNA
You can make your own FERRITE ROD ANTENNA by winding 60 to 80 turns of 0.25mm enamelled wire onto a 9mm ferrite rod or slab. If you wind it on a paper sleeve, you can move the coil along the rod to get the best performance. When the rod is slid out of the coil, the inductance changes considerably. However the inductance does change very slightly when the coil is moved along the rod.

Make your own ferrite antenna

Now we come to the tuning capacitor::

THE TUNING CAPACITOR

The "C" in the "LC" TUNED CIRCUIT can be fixed or variable. When it is variable, it is called a TUNING CAPACITOR. The sheets of aluminium in the air tuning capacitor below are called PLATES and the moving plates are called VANES. The fixed plates make up the STATOR.
The space between the plates is AIR. The photo shows a single capacitor. If two capacitors are connected to the same shaft it is called a GANGED CAPACITOR.
The plates do not come fully out of mesh and that's why the capacitor has a minimum value. The maximum capacitance is when the plates are fully meshed. The odd shape of the plates is designed to produce a fairly constant increase in capacitance as the plates are engaged.
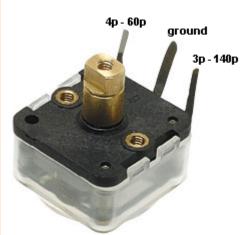
An air tuning capacitor:

Air Tuning Capacitor
(Variable Capacitor)
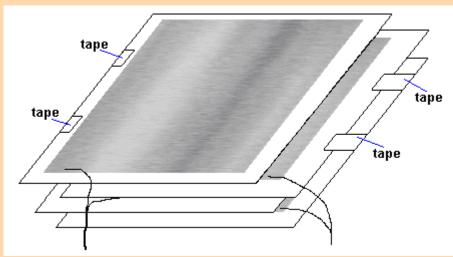
The capacitor can be made much smaller by using thinner vanes and placing plastic between the vanes. Plastic increases the capacitance about 3 times to 10 times.

Tuning Capacitor as found in a pocket radio

The tuning capacitor can be replaced with a home-made equivalent that will work just the same. You need:
4 sheets of aluminium foil (cooking foil) 10cm x 10cm.
4 sheets of thin cardboard 15cm x 20cm   (cut A4 sheets in half).

HOME-MADE CAPACITOR

Tape a sheet of aluminium foil to each sheet of cardboard with sticky-tape around all 4 sides. Take one strand of wire from a length of hook-up flex and sticky-tape the end to each sheet of aluminium to make good contact. Place 2 sheets on top of each other and move the top sheet slightly to the left and sticky-tape the edge so they don't move. Do this with the other two sheets but move the top sheet to the right. Now interleave the sets. Connect the wire from the first sheet to the third sheet. Connect the wire from the second sheet to the fourth sheet.
The cardboard (or paper) between the aluminium sheets increases the capacitance three times. The capacitance decreases when the sheets are moved apart and the capacitance increases when the sheets are moved in. The capacitance also INCREASES when the sheets are squashed together such as when a book is placed on them.
You can also make a smaller capacitor by making each sheet smaller and using 6 sheets. You can then add a 100p or 220p in parallel with the home-made capacitor, to select the lower part of the band.

EACH CIRCUIT

Each circuit we describe in the following set of circuits is an improvement or advancement on the previous. We also offer a number of different types of aerial coils, amplifying stages and earphones. Some of the circuits use easy-to-obtain components and home-made equivalents for hard-to-get items. There will be something in this section for everyone to build.

In all radio circuits you will encounter **TWO MAIN PROBLEMS**:

If the FRONT END (the Coil and Capacitor) is loaded too much by the "pick-off" of the amplifying stages, you will only get one station.

If you get squealing or "motor-boating," try a different circuit and layout as the components you are using, plus the voltage of the battery, will need changing.

You cannot always increase the voltage of the supply and get a louder output. Sometimes the increased voltage will stop the circuit working or it may introduce too much gain that the circuit starts to squeal.

The Radio IC (ZN414) DOES NOT WORK on a voltage above 1.5v and some of the transistor circuits completely stop working with a higher voltage. This has to do with the biasing arrangements and if the circuit is designed for a low voltage, you need to keep to the suggested voltage and experiment with a slight increase in voltage and see what happens.

Building a radio is not easy as the enormous amount of amplification of the combined stages creates a feedback loop via the power rail that sets the circuit into oscillation. This effect gets worse with a higher supply voltage and we will explain this further with each of the circuits.

MAKING A CRYSTAL SET

You can buy a **CRYSTAL SET** kit (see the photo of the kit, above) or the individual components (a kit is the cheapest) or use the replacement for the **FERRITE ANTENNA COIL** (16 turns to 20 turns on a 150mm biscuit tin) and/or the **TUNING CAPACITOR** made from aluminium foil and cardboard sheets.

You will need an outside antenna and an earth (such as a water tap or the frame of your soldering iron) to pick up the radio stations.

If you cannot put up an outside antenna, you will need to add one or more amplifying stages and this will allow you to reduce the length of the antenna and increase the volume of the audio.

ADDING AMPLIFYING STAGES TO A CRYSTAL SET

You can add two different types of amplifying stages to a crystal set.

You can connect amplifying stage(s) to the FRONT END and these will be designed to put less load on the front end so the sensitivity and selectivity increases. These stages work at the frequency of the radio signal and they are called RF STAGES (Radio Frequency Stages).

You can build these stages out of individual components or use a chip called a RADIO CHIP or RADIO IC (integrated circuit) for less than $2.00.

The chip contains 5 stages of amplification and these are RF stages (or RF AMPLIFYING STAGES) and the concept is called TRF. (Tuned Radio Frequency).

It is not easy to get this type of amplifier working because the stages produce a very high overall gain and you get a lot of "motor-boating" and squealing if the gain is not controlled. The gain must be reduced when a strong signal is being passed through the circuit because a strong signal will produce a large output and this will be so large that some of the waveform will find its way to the front of the amplifier via the power rail and start to be amplified again. To prevent the output getting too large, the circuit has a negative feedback line - called the AGC line - Automatic Gain Control.

It would be very difficult to reproduce these 5 stages of amplification with discreet components and that's why it is best to use an IC.

The next stage is a DIODE to convert the RF (Radio Frequency) to AF (Audio Frequency). This can be done with the diode-characteristics of a base-emitter junction in a transistor and we will show the alternatives.

Any stages after the diode are AUDIO STAGES or AUDIO AMPLIFIER STAGES.

The main job of the AUDIO AMPLIFIER is to increase the DRIVE CAPABILITY.

In other words, increase the current capability of the circuit for an 8 ohm speaker or 8ohm earpiece (or 16 or 32 ohm).

This is a very difficult thing to do and requires at least 2 stages.

The LOAD you can put on a Crystal Set must be 10,000 ohms or higher. (if you put a lower resistance (impedance) on the output, you will load the FRONT END and reduce its ability to separate the stations.

That's why a crystal earpiece is normally used with a crystal set. It puts almost NO LOAD on the circuit.

If you put a load on the circuit the result will be only one or two stations across the whole dial and only the most powerful station will be received.

If you don't have a crystal earpiece, you will have to use an 8 ohm earpiece. This will require an IMPEDANCE CONVERTING CIRCUIT of 1,250:1

This is a simple way of saying we want the 8 ohm earpiece to appear as 10,000 ohms to the

crystal set.

To produce an overall gain of 1250, we need two stages of amplification.

If a transistor has a gain of 70, it will it will produce an impedance conversion of 70 times. This is a realistic value. Transistors with a gain of 200 will have a gain of about 70 when fitted to a circuit. This means the other transistor needs to have a gain of about 20 and that is easy to achieve.

ADDING AMPLIFYING STAGES TO THE FRONT OF A CRYSTAL SET

Adding stages to the front of a crystal set are called RF STAGES (Radio Frequency Stages) because they amplify the RADIO STATION SIGNAL.

It does not matter if you amplify RF signals or AF signals. The result is the same.

The only difference is this: The frequency of RF signals is much higher (1,000 times higher) and the coupling capacitors can be much smaller.

This allows an RF amplifier to be built into an IC - called a Radio Chip.
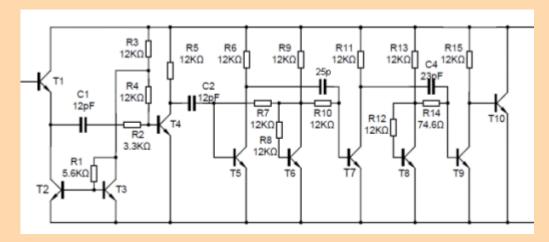
One of the most popular Radio IC's is ZN414 or YS414. This chip has been copied by other manufacturers as: MK484, TA7642 and LMF501T.

All the chips are the same but the pinout is different.
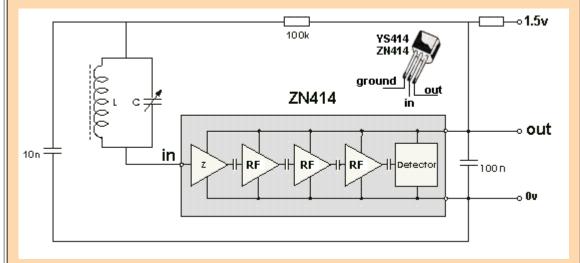
These chips work on a 1.5v supply and if the voltage is increased above 1.5v, the gain of the stages increases to a point of total distortion.

To prevent strong signals producing distortion on 1.5v supply, the output is passed back to the input via a 150k resistor. This feedback line is called the AGC (Automatic Gain Control).

The chip contains 5 stages of amplification plus a stage that converts the RF signal to AF (Detector Stage). This means the signal diode in a Crystal Set is not needed.

Here is the circuit of the TA7642 Radio Chip. It performs the same as the ZN414 Radio Chip.

Here is the BLOCK DIAGRAM of the ZN414 Radio Chip:

The ZN414 chip can be purchased from Talking Electronics for $1.00 plus postage

USING THE ZN414 RADIO IC

By using the ZN414 radio IC (or any if the equivalents) you can create a POCKET RADIO to drive a headphone or speaker.

But it is not easy to use the chip. The main problem is receiving the strong signals without producing distortion and then being able to pick up the weak stations.
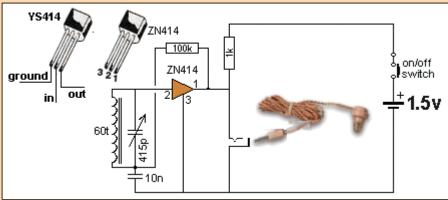
A fixed 100k feedback resistor does not provide adequate control and a TRF radio has limited capabilities.

That's why radio manufacturers make SUPERHETRODYNE receivers. Even though they are more complex, the result is far superior.

However a simple TRF set can be made with the Radio IC and a few stages of audio amplification.

The following circuit uses just the Radio IC and a crystal earpiece or the home-made earpiece described above:

You can use a home-made FRAME ANTENNA or a home-made FERRITE ROD ANTENNA and a home-made VARIABLE CAPACITOR.

The Simplest ZN414 Radio

Connecting the ground (0v rail) to the frame of your soldering iron or a water tap will increase the output volume. The circuit above shows a Crystal Earpiece. Using a Crystal Earpiece may require adding a 10n across the earpiece to improve the output volume. The substitute Piezo Earpiece is effectively a 20n capacitor and an additional capacitor is not needed.

2 TRANSISTOR RADIO

Here is a simple 2-Transistor radio.

The secret to its performance is the 7 turn "pick-off" from the FRONT END (the TUNED CIRCUIT).

The ratio of 7 turns to 60 turns means a small percentage of the voltage generated in the tuned circuit is passed to the transistor. Thus it puts a small load on the TUNED CIRCUIT.

I don't want to go into any mathematics. The turns ratio is 60:7 = 8 but the effect of the 7 turns "pick-off" has an effect called the IMPEDANCE EFFECT and this is the SQUARE OF THE TURNS RATIO. Thus the IMPEDANCE EFFECT is 8 x 8 = 64. This means the "pick-off" (the LOADING EFFECT) is just a few percent. The front end can produce voltages as high as 500mV because a crystal set can produce a voltage high enough to pass through a diode (350mV) and have sufficient to drive a crystal earpiece.
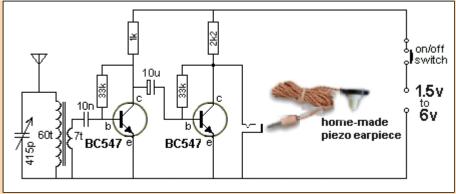
Even though the front end has a "step-down" ratio, the voltage out the 7 turns will be sufficient to drive the first transistor.
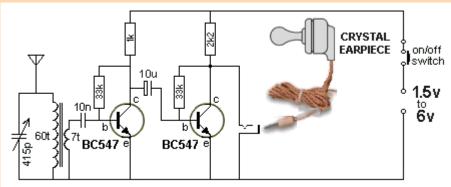
The "transformer" does 2 things: It reduces the loading on the tuned circuit ENORMOUSLY and it produces an output with a higher current than is circuiting in the front end. Even though the transistor is turned ON and biased by the 33k, it is classified as a low-impedance load as far as the front end is concerned and the input signal has to be accompanied by a certain amount of current, otherwise the transistor will not respond to the voltage. The 7-turn "pick-off" is able to provide this current.

Both transistors are biased ON via the 33k base-bias resistors and thus the first transistor responds to the slightest millivolt signal.

This circuit was tested and had the same performance as the **Simplest ZN414 Radio Circuit** above. It can be operated on 1.5v to 6v and the strongest stations tend to overload on 6v.   A short antenna is needed.

**SIMPLEST 2-TRANSISTOR RADIO
using a very-high-impedance earpiece**

**SIMPLEST 2-TRANSISTOR RADIO
using a crystal earpiece**

ADDING AN IMPEDANCE MATCHING STAGE
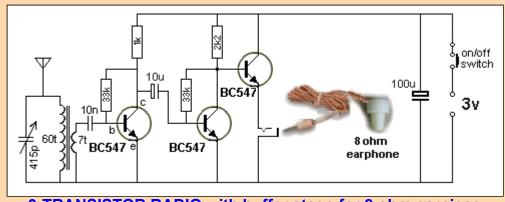
You can add an IMPEDANCE MATCHING STAGE to the output of the circuit above so a low-impedance earpiece can be used.

We call it an IMPEDANCE MATCHING STAGE because this is the correct technical term. It is an AMPLIFYING stage but it amplifies the CURRENT because the second transistor cannot drive an 8 ohm LOAD. 8 ohms is a very low resistance and if it is connected directly to the second transistor, the output will be almost zero.

The reason for this is covered in our discussion: The Transistor Amplifier.

This stage will not increase the volume but simply match the 8 ohm load to the circuit above.

It is very difficult to connect a LOAD to this type of circuit because it will take more current from the battery and cause the supply voltage to fluctuate. These fluctuations will be passed to the first stage and cause variations in the signal. This will be amplified by the first and second transistors in the form of a low-frequency buzzing called **MOTOR-BOATING**.

The only way to reduce or remove this noise is to add an electrolytic across the power rails and reduce the supply voltage. The third transistor simply takes the waveform on the output of the second transistor and delivers it to the earphone with a higher current. It is called an IMPEDANCE MATCHING STAGE as it effectively increases the 8 ohm load by a factor of about 100.

3-TRANSISTOR RADIO with buffer stage for 8 ohm earpiece

The 3rd transistor converts the 8R to about 800R

You can use 16 ohm, 32 ohm or 64 ohm in place of the 8R earpiece and these will give better performance as they will take less current and improve the stability of the circuit.
Low-impedance earphones create "motor-boating" due to the peaks of current and this can be very hard to fix.

A 2-TRANSISTOR RADIO with REGENERATION

The next stage in our discussion to get better performance is a feature called **REGENERATION**.
Regeneration sends a small output signal back to a previous stage in the form of POSITIVE FEEDBACK to INCREASE the original signal. The signal on the emitter of the first transistor is the same amplitude as the signal entering the base but the FRAME ANTENNA has a turns ratio of 5:15 and this increases the signal on the receiving section of the antenna by up to 3 times. But we want the returning signal to be just above the amplitude of the receiving signal and so a resistive adjustment (attenuator) is provided on the emitter to deliver just the right amplitude.
This has the effect of increasing the amplitude on the TUNED CIRCUIT and is just like reducing the load on the circuit.
As we have mentioned above, when the tuned circuit is lightly loaded, it will pick up a station at the exact frequency of transmission and if the dial is changed slightly, the station will disappear.
This quality is called SELECTIVITY.
At the same time, the Tuned Circuit will pick up weak stations and this is called SENSITIVITY.
The quality of a receiver depends on the loading of the TUNED CIRCUIT.
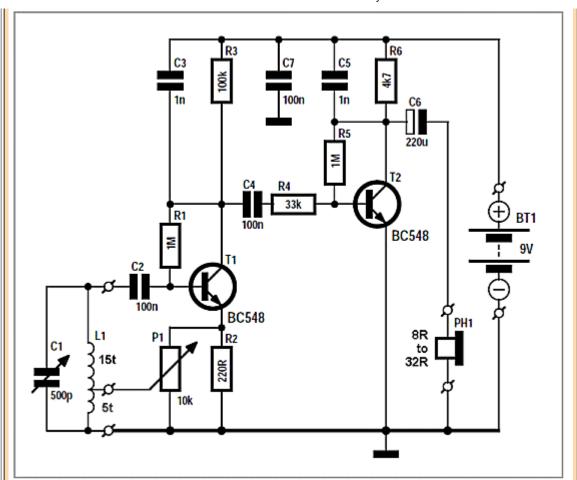
Here is the original circuit from Elektor Magazine with the prototype made on matrix board and fixed to a base-board with a frame antenna made from two sticks of wood. The photo shows a speaker but the output is so low that you really need headphones.
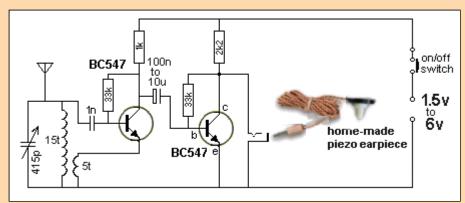
2 TRANSISTOR REGENERATIVE RADIO from Elektor Magazine

The circuit is very complex and the output will be very low as the circuit cannot drive a low-impedance earphone via a 4k7 load resistor. The 4k7 resistor is actually driving the speaker (the transistor is simply discharging the 220u). The 4k7 only allows 32/4700 x 9 = 61mV to appear across the earphone - a very poor result.
The skill of designing a transistor stage is covered in our comprehensive eBook: The Transistor Amplifier and you wont make a mistake like this !!!
The circuit above can be simplified and we can add the REGENERATIVE feature to our **Simplest 2-Transistor Radio** circuit:

Our circuit uses a 15 turn circular FRAME ANTENNA 15cm diameter and a 5 turn REGENERATION coil.

2 TRANSISTOR RADIO with REGENERATION

The regeneration coil is brought near the main coil and as it gets closer you can hear the audio get louder. If this does not happen, turn the coil around.
Early radios used this technique and the operator had to adjust the coil by hand. No-one minded because radio was a fascination and the simplest radio cost more than a weeks wages. To listen
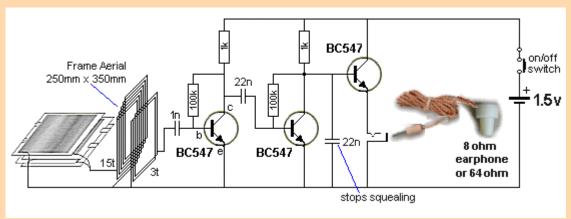
to a broadcast through headphones was an amazement and listeners would sit all night with headphones listening to music.

This is very fiddly and by adding an extra buffer stage, we can use a Frame Antenna with a very clever "pick-off" that does not load the front end. This gives the circuit very good sensitivity and selectivity without regeneration.

3 TRANSISTOR RADIO

Here is our final design for the simplest self-contained 3-Transistor Radio using our home-made Tuning Capacitor and 250mm x 350mm Frame Antenna. It picks up the local stations and drives a low-impedance earphone or set of earphones (from a mobile phone).

3-TRANSISTOR RADIO

The circuit performs very well and uses readily-available components. The 22n across the output is essential to stop squealing.

The secret to sensitivity and selectivity is the turns-ratio on the Frame Antenna. The 3-turn "pick-off" puts very little load on the front end and this allows the stations to be tuned with our home-made Tuning Capacitor.

The circuit contains all the features we have discussed above and only needs a 1.5v supply. Build this circuit before you buy any expensive tuning capacitors, IC's or ferrite slab antennas as you will not get any better results.

This is called a TRF circuit and because the stages operate at Radio Frequency or Audio Frequency. Due to the high amount of amplification, the circuit can start to squeal (feedback, motorboat) due to the layout.

You may need to shorten or lengthen the leads or move the parts slightly - it's that critical. However the result is a portable radio that needs no earth and will pick up the strong stations. You can try connecting the 0v rail to the metal part of a soldering iron to increase the number of stations.

LOADING

The whole success of picking up a radio station is the RECEIVING CIRCUIT. The receiving circuit is the coil and the signal in the air (from the radio station) must go down the centre of the coil.

It cannot pass over the top or the bottom of the coil. Only the signal that goes down the centre of the coil is received.

As you can see, the centre of the coil is not very big and it is amazing that the signal can pass down the centre. But it does, and that is the only signal that will be amplified.

This signal is passed to the capacitor and we have explained how the signal is gradually increased and increased in amplitude until it is as large as 500mV. The signal from the radio station may be as small as a few millivolts, but as it keeps pushing the "swing" back and forth, the amplitude get larger and larger.

If you put your finger on the "swing" you will prevent it get larger and larger and it only requires the slightest touch of your finger to prevent the swing gaining full amplitude.

In electronic terms, your finger is called LOADING THE CIRCUIT and since we have to pass the signal to further stages of amplification, we need to "tap" or "load" or "pick-off" a signal.

The aim is to load the circuit as least as possible because the actual energy entering the circuit is very small.

In fact, this is all the energy we can remove as that is all the energy entering it.

Because a very small amount of energy is entering the "front-end" we classify it having a very

high impedance. It is very difficult to provide a value of impedance for this circuit because impedance has the term "Z" and the circuit is operating a very high frequency so resistance values are not the same as impedance values.

The actual resistance of the circuit is ONE OHM but the impedance is more like 10,000 ohms to 100,000 ohms.

We can explain its high impedance if we put a 100,000 ohm resistor across the circuit. The waveform will be reduced very slightly. If we put a 10,000 ohm resistor across the circuit, the signal will be reduced a reasonably large amount. If we put 1,000 ohms across the circuit it will stop working.

This means a load of 100,000 ohms will have the least effect.

In a crystal set, the diode creates NO LOAD until a voltage of 350mV is reached. It then passes excess voltage to a crystal earpiece that has a very high impedance. That's why a crystal set will produce a good output. The LOADING is very small.

When a transistor is connected to the TUNED CIRCUIT, it starts to put a load on the circuit after 600mV and this load is VERY HIGH. The "resistance" of the base-emitter junction is about 1k and the signal will find it very difficult to rise above 600mV because the incoming energy is not sufficient to increase the voltage.

Adding a capacitor between the base and the front end allows the transistor to be self-biased and get a turn-on voltage of about 600mV from a base-bias resistor.

The FRONT END is now separated from the transistor and ANY voltage it is producing will be passed to the transistor via the capacitor.

Whereas, with the crystal set, the first 350mV could be produced without any loading, the circuit is now loaded AT ALL TIMES.

This means we have to load the circuit as lightly as possible to be able to pick up individual stations.

The only way we can do this is to use a capacitor of the smallest practical value and this has to be worked out by trying different values. If the value is too small, the transistor will not detect a small signal. If the value is too large, the circuit will stop working.

Values such as 1n, 10n and 100n are suitable.

Values such as 1u, or 10u will be too large.

The CRYSTAL SET loading and a transistor load are completely different.

The transistor loads the front end ALL THE TIME and that's why you need to use a transformer or other ways to reduce the loading. Sometimes a Field Effect Transistor is used as it puts almost no load on the front end.

2-4-2016
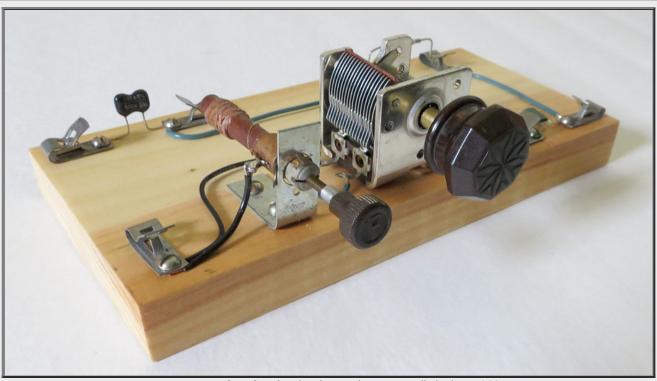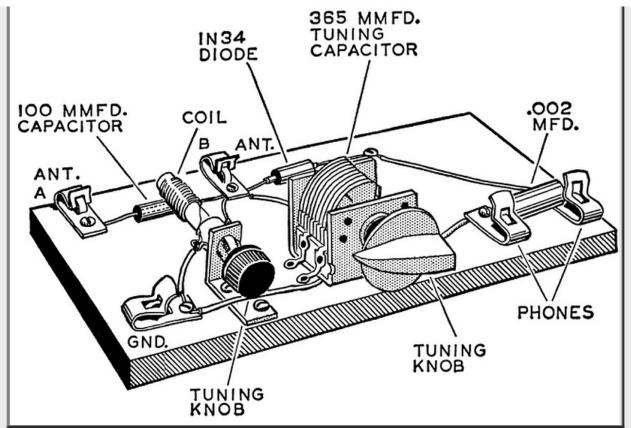
Recreating an Alfred P. Morgan crystal set

A recreation of my first (working) crystal set, originally built in 1966.

Page from "The Boys' Third Book of Radio and Electronics" by Alfred P. Morgan.

Back Story:

In the Summer of 1966 I was 10 1/2 years old and one of my favorite TV shows was "Get Smart". It was a show about a bumbling secret agent named Maxwell Smart who had all sorts of gadgets at his disposal. My friend Billy Meyers and I decided we wanted to be secret agents, like Max. One of the "secret agent" things we'd do was to pick out a guy with a briefcase walking home from work and declare him a Soviet spy. Then we would "tail" him for a few blocks, making up stories about him.

Billy and I had no way to communicate with each other after we had to come in for the night. He lived half a block away. Kids didn't use the phone back then. There was only one phone in the house and I don't remember using it before I was about 14 years old. Maxwell Smart had a phone in the sole of his shoe. We needed something like that!

Maxwell Smart (Don Adams) Agent 86.

Remco "Monkey Division" Wrist Radios.

One thing I DID have was a set of Monkey Division Wrist Radios. These were powered by a single C battery in the "Master" unit, which also had a button on it that would buzz the other receiver to get the users attention (apparently, the other user was unworthy of a button). A metallic speaker doubled as a microphone. The sound was very tinny, but you could make it out. Unfortunately, they were wired to each other. There was nothing "radio" about them.

I remember taking them outside with my brother Rob, and I could see and hear him talking at the same time his voice was coming over the wrist radio. They were pretty much useless, unless you like running around with a wire connecting you to your brother.

BUT... with some extra wire strung across the driveway, down the backs of the houses and into Billy's bedroom, Billy and I would be able to talk to each other! I immediately presented this great idea to my mom with a request the she fund the cost of the wire, and she immediately refused and told me the electric company would just come out and take the wire down.

Location of my house and Billy's house (actually his grandparent's house) The wire would have to be strung across the driveway. This is part of the West Oak Lane section of Philadelphia.
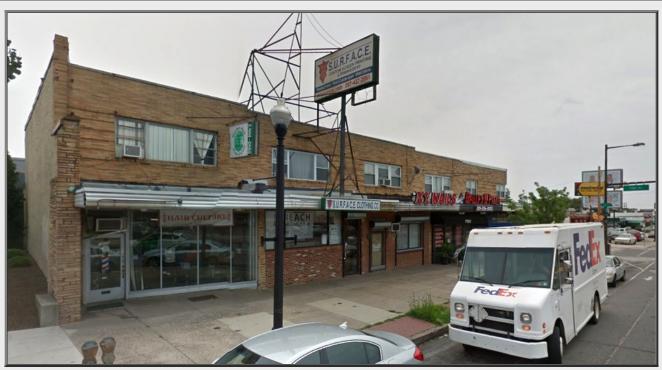
There was only one thing left to do. Build a radio. I went down the basement and connected a battery to a speaker and used a coat

hanger as an antenna. This was similar to the wrist radio, but the coat hanger antenna replaced the wire. It didn't work! All it did was make clicking sounds in the speaker. I had to wait for my dad to come home from work and ask him why.
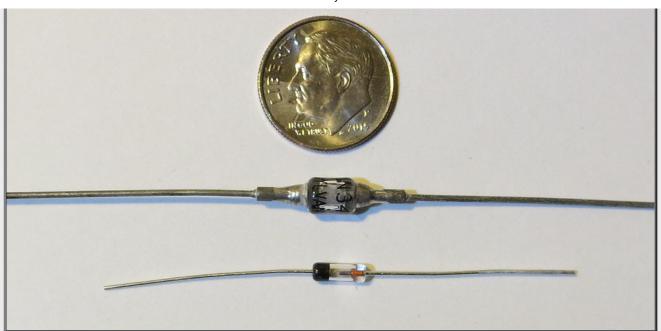
When my dad came home I showed him the setup and he said, "You don't have a detector." I asked him what a detector was and he told me to go the library and get a book on radio. The next day I had a copy of "**The Boys' Second Book of Radio and Electronics**" by Alfred P. Morgan. Chapter 2, page 15 was titled "Building Your First Radio Receiver". I wouldn't be able to talk to Billy with it, but that was OK. My mom wouldn't buy me a secret agent coat, we didn't have any gadgets, and our secret agent days were coming to an end.

Now there was another problem. None of the parts needed to build anything in the book could be found around my dad's workbench. There was a store on Ogontz Avenue named REE Electronics, so I headed up there with a list of parts. The store sold stereo equipment and fortunately for me, also repaired stereo equipment. I asked the man in the store if he sold diodes or "capacitaters" and he sent me into the back of the place. There were two guys back there and bins of parts along the wall.

REE Electronics was located at 7709 - 7711 Ogontz Avenue in Philadelphia. The entire block has been razed and rebuilt, and is no longer recognizable. The picture above is the 7900 block of Ogontz Avenue. The store on the left is the only one that retains its original appearance, with the glass store window and the apartment overhead. This is how REE Electronics looked in 1966.

The two guys were pretty cool. I announced that I would like a "three hundred and sixty five micro micro farad variable capacitater". They asked me a couple of questions and told me to come back with the book. They had all the parts I needed except the coil. No problem, I would just build the set with no coil. I came home with Fahnestock clips, a 1N34 Germanium diode, a variable capacitor and a crystal earplug.
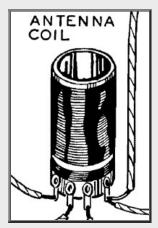
A Sylvania 1N34A Germanium diode (top) from 1949 and its modern counterpart. The new diodes are literally a dime a dozen, less than one cent each. The 1949 Sylvania diode cost me $9.00 in 2015. It is a duplicate of the one I bought at REE Electronics in 1966.

The diode was 65 cents in 1966. I can't use the $9.00 diode because I don't want to bend the leads, so I sort of just look at it.

Of course the radio didn't work without a coil. It did pick up the slightest whisper of KYW AM 1060 mingled with WIBG AM 990. There seemed to be some buzzing associated with it, as what I could hear sounded distorted. I HEARD something, that was the really, really neat part. It made such an impression on me that I remember the date. July 26, 1966.

The "problem" with the Alfred P. Morgan books was that they were not written for anybody as dumb as I was. Morgan didn't write, "If you can't find the coil you can make one." He just said to go buy a coil. Not only that, but there were no photographs in the book, though there were excellent drawings on almost every page. Since I had never seen some of the parts in real life, I didn't know exactly what the coil looked like, and I didn't understand what it did. That's because I didn't read the book! I was stuck on page 15.
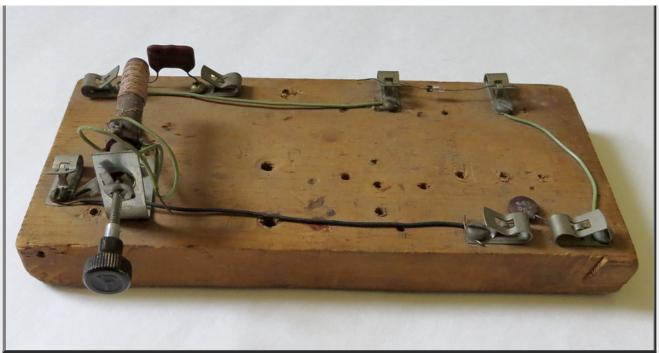
This is the picture of the coil from the book. I didn't know what I was looking at.

I returned the book to the library and came home with "**The Boys' Third Book of Radio and Electronics**." I found a simple radio on page 104 and soon headed back to REE Electronics. This time, they DID have the coil! I can't remember how much time passed after the first non-functioning radio was built. Probably a month or so.
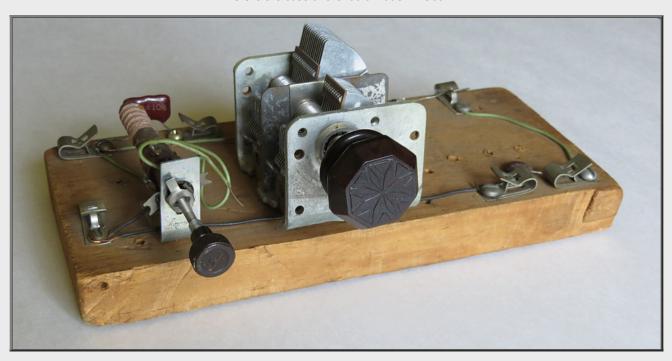
I asked my dad to cut me a wooden base for the radio. I started building the radio and, if I remember correctly, it took me a long time. I didn't have a drill, so any holes in the base were made with the point of a compass. There were three connections that needed soldering. I got some solder from the basement, and the tweezers and alcohol lamp from my chemistry set. The tweezers were heated in the flame of the lamp till the tips began to glow, then I would quickly solder the joint.

One day a friend from school named Leo Pound stopped by on his bicycle. This was a bit unusual because Leo lived miles away. I don't even know how he knew where I lived. He recently told me (via Facebook) that he remembers helping me build the radio. Odd that it was the one and only time he came by. Apparently, we got the radio working that very day.

Here is the base of the radio made in 1966.

I found all the original parts except for the tuning capacitor. However, it looked like this one. It's just sitting on the base in this photo, but the radio pretty much looked exactly like this. My dad gave me the big tuning knob. The smaller knob came from a lamp in my bedroom. I didn't tell my mom you could no longer turn on the lamp, but eventually I found a lamp in somebody's trash and took the knob off to replace the one on my lamp.

Original coil, diode and main tuning knob from the 1966 radio. The wires on the coil were soldered with a pair of red hot tweezers.

My dad gave me the large knob for the tuning capacitor. He had a second job on the weekends at "John Cusimina's Moving and Storage." I've always wondered if he pulled that knob off of somebody's radio while they were moving. I hope not. He probably did.

Next

CRYSTAL SETS 5
Experimental Crystal Sets
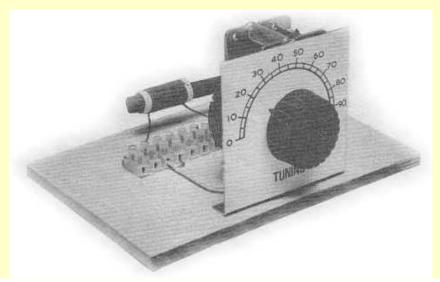
CRYSTAL SETS 5: EXPERIMENTAL CRYSTAL SETS
.

Picture 1 - The Complete Experimental Crystal Set

THE POPULARITY of the crystal radio arises from its simplicity, and the fact that it needs no power supply. The circuit here allows for easy experiments with tuning, aerial and diode coupling, and frequency coverage. Wrong connections can cause no damage to any components.

A Crystal Set is more often than not used for the reception of medium and long wave radio, but short wave reception is also quite feasible. It will normally be possible to receive some of the stronger international radio stations.

This is adapted from an article that appeared in the 1970's in Everyday Electronics, and gave me almost endless hours of fun!

BASIC CIRCUIT

The basic circuit is shown in Picture 2 below. The coil L1 can be air cored, or have a ferrite rod placed in its winding. The variable capacitor C1, in conjunction with aerial-earth capacitance, tunes the circuit to resonate with the wanted radio station frequency. The diode D1 "detects" or demodulates the radio signal so that the programme is heard in the earpiece.

This basic circuit can be modified in various ways to obtain better performance.

EARPHONE

As most constructors will be using a Crystal Earpice to listen to the crystal set it is essential that a 47k Ohm resistor is connected across the earphone terminals (TB1/1 and TB1/2 in the diagram), i.e. in parallel with the earphone, otherwise results will be very quiet.

A High Impedance headset of 20k Ohms (20,000 Ohms) may give even better results, but these are very difficult to obtain , so unless you happen to already own such a headset the Crystal Earphone with 47k resistor will be the only option. An ordinary magnetic earpiece or Walkman headphones will not work with a crystal set.

ASSEMBLY

Construction is of a 'breadboard' type using a wooden board of about 165 x 130 mm. A 12-way block connector, TB1, is used to connected together the components and this is screwed onto the wooden   board. The use of a block connector provides an easy method of connecting the components together and then subsequently rearranging them as the experiments progress.

Tuning capacitor C1 is screwed to a bracket made of some scrap metal which is then also screwed firmly down to the baseboard, see Picture 1 above. Thin plywood screwed to the front edge of the baseboard would also provide a suitable method of fixing the tuning capacitor to the base. A knob with pointer is fitted to C1, and a scale is drawn and fitted behind this.

Except for C1, all connections are made by the terminals of the 12-way terminal block as shown in Picture 4. Loosen the screws with a small screwdriver, insert the bared ends of the wires, and tighten the screws. The various locations on the terminal block, TB1, are also shown in the circuit diagram, Picture 2.
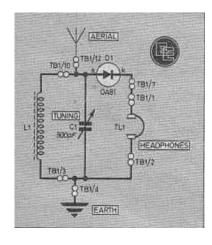
AERIAL AND EARTH

Crystal receivers need a long wire aerial preferably strung outside and about 25m long, or as long as is possible to install. If this is outside it should be high and clear of earthed objects as this will improve performance.
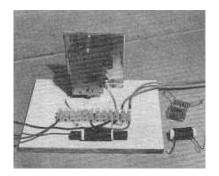
An earth is absolutely essential for a crystal set to work properly. The earth lead can be run to an earth rod or spike that is buried to a depth of about 1 meter into damp soil. Or it may be soldered to a bare metal can which is buried in damp soil.

It is feasible, though not recommended, that the earth lead can be connected to the earthing terminal of a hi-fi system or even to the bare metal case of a personal computer that is plugged into an earthed mains outlet, but is switched OFF.

Stranded, insulated wire, or purpose made aerial wire can be used for the aerial and earth leads.

Picture 2 - The Basic Circuit

Picture 3 - Photo Of The General Layout

INDUCTORS (The Tuning Coils)

The following four coils are suggested for initial use as L1 :

Coil 1: Make a thin card tube to slide on a 10mm diameter ferrite rod, and on this tube wind about 105 turns of 32 s.w.g. enamelled copper wire, side by side. Secure ends with sticky tape.

Coil 2: Make a similar coil to to coil 1 having about 15 turns of 24 s.w.g. enamelled wire on the card tube. Loops of cotton will help hold the ends in place.

Coil 3: Wind 9 turns of 20 s.w.g. bare tinned copper wire on an object about 20mm in diameter. Remove and stretch to separate the turns, to obtain a coil about 25mm long.

Coil 4: Make a similar coil to coil 3, but with 5 turns.

The Ferrite Rod

It will be necessary to have a ferrite rod of about 60mm to 75mm long available. Coils 1 and 2 will provide reception of medium wave and the longer short wave bands. Coil 3 should cover about 3 - 10MHz shortwave with the ferrite placed in it, or about 6 - 18MHz with the ferrite rod removed. Coil 4 should cover about 6 -13MHz with the rod in, and about 9 - 20MHz without the' rod.
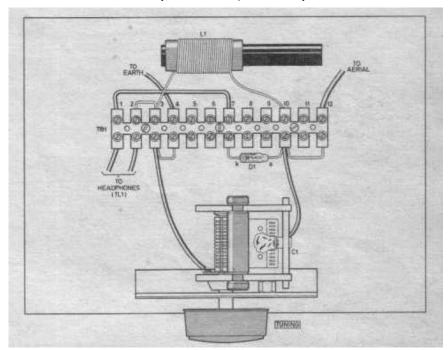
It will be noted that as the ferrite rod is inserted, any particular signal has to be re-tuned by opening Cl. This arises because the ferrite increases the inductance of the winding, so less parallel capacitance is needed for the same resonant frequency.

EFFICIENCY CHECKS

Tune in a m.w. transmission using coil 1 which gives good headphone volume. Place a microammeter or multi-range meter on a sensitive range in series with the headphones. A reading of 50-100uA or more may be obtained, depending on aerial, earth, earphone resistance and resistor value, coil and detector efficiency and strength of signals at your locality.

Placing the ferrite rod in the coil and re-tuning should boost the meter reading to some extent. Surplus or other detector diodes can be tried by substituting them in turn and noting the meter reading. Improvements to the aerial (or earth) will also show up as a rise in meter reading.

If experimenting with a crystal earpiece, which gives no direct current circuit, the meter may be clipped across the phone leads, i.e. D1 cathode to earth.

Picture 4 - Baseboard Layout Of The Crystal Set

AERIAL COUPLING

The aerial loads the tuned circuit heavily when connected directly to the top of the tuned circuit, as in Picture 2. This damps the tuning action and it can be found that stations spread out all over the dial, which is unsatisfactory.

The series capacitor, C2 connected in Picture 5(a) reduces the loading and thus improves the sharpness of the tuning. A variable or pre-set capacitor of about 250pF maximum is most suitable. for this role, though it is possible to experiment with a variety of fixed value capacitors in this range also.
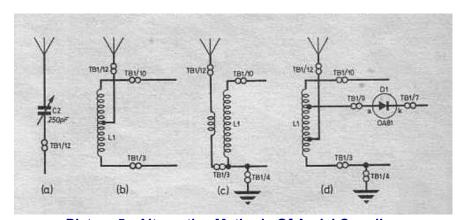
Connecting the aerial to a tapping on the coil, as in Picture 5 (b) also sharpens tuning. It may also increase volume. Try about 2 turns from earth for coil 4, or 4 turns from earth for coil 3.

Another method is to have a coupling primary, as in Picture 5 (c). This consists of a second coil, with about one third the turns of the original wound on top of the existing coil.

You can even combine these methods to find what arrangement best suits the aerial in use.

The diode can be disconnected from the end of L1 and taken to a spare position on TB1 for example location TB1/9. You can then run a flying-lead fitted with a crocodile clip from this position, connecting it to various tappings on the coil as required as in Picture 5 (d). This method also reduces loading on the tuned circuit.

Coils with spaced turns of bare wire are readily tapped. For other coils, small loops can be made every ten turns or so, and crocodile clips can be attached to these when selecting tapping points.

Picture 5 - Alternative Methods Of Aerial Coupling

SHORT WAVES

For shortwave reception, a good efficient outdoor aerial is certainly recommended. Evening listening in the region around 5 - 9MHz in often proves to be the most fruitful.
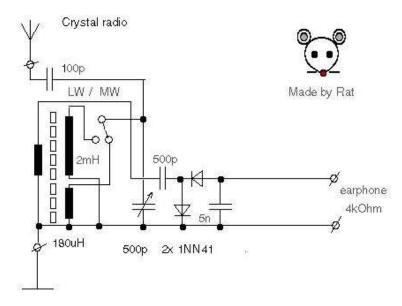
Since there is no amplification, as with a valve or transistor receiver, certain frequencies will seem to be completely dead at particular times of day. So if the crystal receiver works satisfactorily on medium wave and longwave, but no shortwave signals are heard, check again in the evening, or after dark, when conditions are different.

PARTS REQUIRED
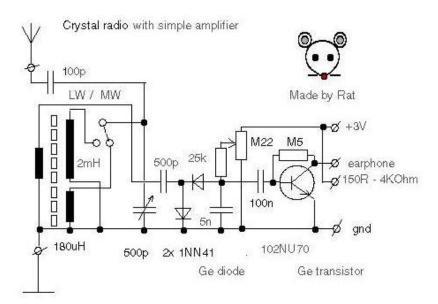
C1	365pF or 500pF Air Spaced Tuning Capacitor
D1*	**OA47, IN34, OA81, OA90, OA91, IN94 or similar point contact small signal Germanium Diode** * The OA47 will be of particular interest since it has the lowest forward bias voltage of any of these diodes which will make the crystal set somewhat more sensitive and therefore louder. The US equivalent of the British OA47 is the IN34.
TL1	High Impedance Headphones (20,000 Ohms) or Crystal Earphone
TB1	12-Way Plastic Screw Block Terminal
Also Required:	47 k Ohm Resistor for Crystal Earphone: Enamelled Copper Wire: 32 and 24 s.w.g. for L1: 20 s.w.g. tinned wire for L1: Ferrite Rod 10mm diameter x 75 mm long: 25m of wire for aerial: Wire and rod or spike etc for earth: Wood for base e.g. 10mm x165mm x 130mm: Scrap of metal of thin plywood for C1 bracket/front panel: Knob: Crocodile clip(s)

Adapted from an article in Everyday Electronics magazine, November 1981, By F.G. Rayer.
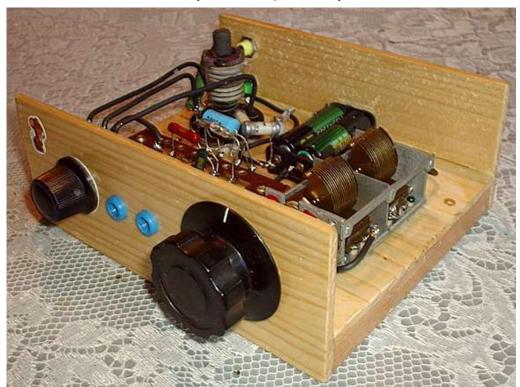
**HERE ARE A COUPLE OF VERY INTERESTING CRYSTAL SET DESIGNS
SENT IN BY KRYSATEC - "THE RAT" - FROM THE CZECH REPUBLIC**

Crystal radio

Made by Rat

100p

LW / MW

2mH

500p

180uH

500p   2x 1NN41

5n

earphone
4kOhm

1/ Using old coils from old bulb radio for MW and LW band. Though it would be straightforward to wind the coils - one for Long Wave, one for Medium Wave and a coupling coil. Variable capacitor is 2 x 500pF only one half is used: 500pF. For the crystal earphone a resistor of about 82k ohm in parallel is required. This set also uses two Ge diodes as a multiplier in the quest for for higher audio signal output.

Crystal radio with simple amplifier

Made by Rat

100p

LW / MW

2mH

500p   25k   M22   M5

180uH

500p   2x 1NN41

5n

100n

102NU70

Ge diode   Ge transistor

+3V

earphone
150R - 4KOhm

gnd

2. If signals are not strong signal in your location, then the above circuit design can be considered. A simple transistor amplifier is used. A variable resistor M22 is used for better sensitivity which can be adjusted for poor signals. This crystal radio is aversion from cca 1960 - 1970 y.

Rat's finished Crystal Set with additional amplification - very neat!

'Minilabs' Crystal Radio

BELOW: Ian Tomlinson kindly sent in a photograph of the box that contained the kit for his John Adams Toys 'Minilabs' Crystal Radio.

It is a very simple circuit consisting of the coil (inductor) with a sliding contact that provides variable tapping points, a diode and crystal earphone. All that is added is the aerial and earth. There is no variable tuning capacitor for simplicity and to keep costs down.

The coil provides the inductance required for tuning into a certain frequency (wavelength). These days a variable "tuning" capacitor is normally wired in parallel across the inductance (coil) in order to vary the resonance of the tuned circuit and therefore enable to easily tune into various transmitters on different frequencies. This crystal is tuned varying the number of turns on the coil (ie varying the inductance) by tapping off at different points using the sliding contact ("ball").

The crystal earpiece, or high Z headphone, is connected between the output of the detector diode (the other end from the coil) and earth. The volume from a crystal earpiece may be considerably improved by connecting a resistor of - somewhere between - 4.7 k and 47k ohms in parallel with the earpiece. A crystal earpiece cannot directly allow current to flow through it and the parallel resistor therefore allows current to better flow through the circuit.

'Minilabs' crystal set by John Adams Toys

A discussion on configurations for Crystal Sets by Felix Scerri VK4FUQ

This discussion, by Felix Scerri VK4FUQ, was posted at this address which no longer appears on the web www.tarc.org.au/techinfo2.htm (error 404) so here it is reproduced:

Crystal Set design is one of my passions closely allied with my obsession for audio and high fidelity.

My main interest in crystal sets, apart from the wonder of a radio receiver that does not require a power source, is the potential excellence of the recovered audio quality from normal AM broadcast stations.

Personally, it is one of my great laments that most people have never heard how good wideband AM can sound. A high performance crystal set or similar TRF approach is, in my opinion,the only way to do it. There are a few people around who have heard the audible results of my efforts,and can only agree.

I have often wondered,given the ultimate simplicity of the crystal set, being essentially a tuned circuit,a diode detector and some form of output device, what it takes to achieve optimum performance. What follows are my thoughts on the matter.

Crystal Set optimisation, is in my opinion, all about reduction of circuit losses. Essentially this means high "Q" tuned circuits and high quality detectors. Efficient output devices also help too. But as we will see, there are some trade-offs required as well. A high "Q" tuned circuit is always benefical, as a high "Q" tuned circuit has lowest RF losses,highest potential selectivity,and highest voltage at resonance, which is very useful for the diode being fed from the tuned circuit. Variable capacitors, even the "modern" miniature variable capacitors (although the older air dielectric units, as used in old valve receivers are more desirable) for various reasons,are generally quite efficient, and a higher "Q" coil will produce the most worthwhile improvements.The best (highest "Q") coils are wound with "Litz" wire, which is a multistranded woven wire with all strands insulated from each other. The performance of Litz wire wound coils is spectacular, unfortunately, although I know Litz wire is still being made, from personal

experience, it is VERY rare in Australia.

Efficient coil design can be quite complex and all my coils are wound on ferrite rods. There seems to be,at least for ordinary single wire windings (close wound), an optimum wire thickness for optimum coil "Q". I have determined .315 mm winding wire to be about optimum for simple (single wire) coils on ferrite rods. Thicker wire is NOT better, believe it or not.

Lacking Litz wire, an interesting winding approach I have developed is to use two slightly thinner wires wound as a bifilar winding connected together at the beginning and end of the coil, yields considerably higher "Q" compared to a simple single wire winding. I have found 0.25 mm winding wire optimum in this application.

Whilst high "Q" coils are beneficial from the RF point of view, there is a possible downside. If one is interested in maximum selectivity and sensitivity, there is no problem, but remember highest "Q" results in a narrowed audio band-width as a simple consequence of band-width. For high fidelity applications this could be a disadvantage under some circumstances, although there are clever ways around this.

Regardless of ultimate coil "Q", selectivity is a major issue with crystal sets generally. Here another trade-off is evident. For the maximum voltage into the diode, connecting the diode to the high impedance end of the coil (i.e. the top) yields the greatest voltage but the selectivity is usually terrible, because of severe "loading" by the diode circuit. For this reason, tapping well own the coil improves selectivity at the expense of signal volume (reduced voltage). Once again there are ways around this. As described in my "Double Tuned Crystal Set Tuner" article in "Amateur Radio" magazine, March 2002, the use of two separately tuned coupled resonant circuits allows top connection into the diode without compromising overall selectivity, thanks to the use of a second tuned circuit which is fed from the external antenna. The whole network forms a double tuned input bandpass filter and in practice this approach works very well. For single coil crystal sets I recommend the use of an un-tuned "antenna" winding adjacent to the "hot" end of the main coil, preferably adjustable (old paper reels from sewing cotton threads are ideal). This allows the degree of coupling to be optimised under actual listening conditions. The double tuned set up is best, yielding superb selectivity, but the un-tuned antenna coil arrangement also works quite well, especially if the diode is tapped well down the main coil.Tapping halfway works well.

The other method of performance improvement involves the use of the most effective detector system possible. Here things get very interesting. In fact the temptation is to use more complex circuitry, but that gets away from the charming simplicity of the crystal set. As an example, my own crystal set tuner has at times mutated into a TRF tuner complete with FET RF preamplifiers, active(powered) detectors and other enhancements. These modifications do work well, but loses the simplicity of a basic crystal set. In actuality, a simple diode detector can work extremely well, subject to some qualification. Diodes like to work with a reasonable level of RF input voltage. Audio distortion can result under conditions of low signal level, due to diode transfer curve non linearity and other factors, such as the widespread use of broadcast station "processing". The actual type of diode makes a difference. The 1N34A germanium diode is very popular for crystal set use, although in my experience just about ANY germanium diode will work, although it is worth trying different specimens. Some are definitely better than others. Even from a pack of twenty 1N34A's from the same source, some were definitely better than others. Measuring the average value of rectified output voltage across the diode load resistor will show which diodes are best. By the way, I regard a diode load resistor as being mandatory. I find a value of about 47K about right, especially if a crystal earpiece is being used or the crystal set is being used as a tuner feeding an audio pre-amplifier and following amplifier. If using high impedance magnetic type headphones, the headphones provide the diode DC load.

Another type of diode that is very interesting, is the hot carrier diode. There seem to be a lot of different hot carrier diodes around these days. There are even hot carrier diodes now being sold as "germanium diode equivalents". I have tried them and they do work acceptably well, but they are not quite as good as genuine germanium diodes such as the 1N34A. Typical UHF mixer hot carrier diodes, such as the 1N5711 will not work well in crystal set service simply because their "turn on voltage"is too high, similar to silicon diodes such as the 1N4148/914 series, which require a lot of RF input to function adequately as RF detectors, however a simple technique can be used to turn hot carrier diodes such as the 1N5711 into superlative detectors.

I guess we are cheating a little, because the technique is to use a little voltage bias supplied via a 1.5v battery, through a simple potentiometer voltage divider arrangement, with capacitor (for DC isolation) fed into the diode from the tuned circuit. With applied adjustable bias, I find the 1N5711 diodes absolutely superlative detectors under ANY signal strength conditions. I find the detection quality also superlative, with a clarity and low noise profile unmatched by any other diode arrangement. In my opinion, hot carrier diodes, running bias,are the best detectors overall.
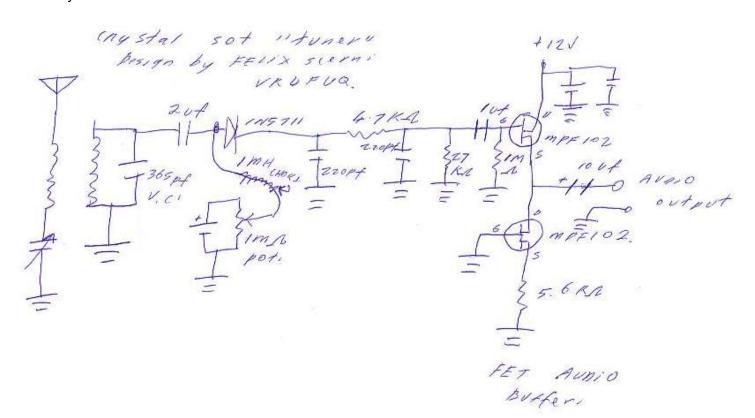
Regarding other detector arrangements, the diode "voltage doubler" is often recommended, however my own

experiments with the doubler arrangement have been inconclusive and slightly disappointing overall. I have found no real advantage in their use over a simple (one) diode detector, believe it or not.

Yes, they do work, but they're nothing special, at least in my opinion.

Any comments on this general subject of crystal set optimisation would be welcome.

73's Felix Scerri VK4FUQ.
22nd July 2002

Above: CRYSTAL SET BASED CIRCUIT PROVIDING A HIGH QUALITY PROGRAMME SOURCE

IMPROVED VERSION OF THE ABOVE CONCEPT !! New update from Felix Scerri February 2010:

New 'two FET infinite impedance AM detector'

I've developed a new version of my old favourite FET 'infinite impedance' AM detector that I think sounds very nice. I include a short audio of one of our local AM stations. I picked this station as it is my reference 'torture test AM station' as they run very heavy 'processing' which normally sounds yuck with all my other (diode and non diode) detectors! However it's quite clean with this detector. What do you reckon? I'll do up a circuit if you'd like to feature it in your TRF radio section. A general draft article follows.

'A favourite non diode based AM detector that I've built and used many times over the years is the FET based infinite impedance detector, offering very good general AM detector performance, especially under weak RF signal conditions where diode based detectors do not perform well, especially in terms of audio distortion.

However one of the slightly strange things I've noticed about the simple FET based infinite impedance detector is the variable audio quality noted, even when using the same type of FET. Some I've built have sounded good and others slightly fuzzy when used with an audio preamp and fed into a high quality audio system. I've been giving this a considerable bit of thought of late and I've wondered if the audio distortion might be a result not necessarily of the detection process itself, but the FET stage in its guise as a 'source follower' audio stage which essentially, it is.

I have long been aware that as a simple audio buffer stage, the FET based 'source follower' can exhibit a considerable amount of audio distortion, and a technique I've long used to greatly reduce this audio distortion is to use a second FET in the source lead of the first FET as a 'constant current source' which serves to 'linearise' and

greatly reduce audio distortion in the buffer stage overall. So, to test the theory I built a simple one FET infinite impedance AM detector which worked well, but with just a hint of audio 'fuzziness' on received AM stations. So I added a second FET in the source lead of the first FET wired as a constant current source, taking the output from the source of the first RF detector FET and the source resistor and RF bypass capacitor off the source lead of the second FET 'constant current source'. The result, totally clean audio! The theory seems proved! I call this modified detector the 'Two FET infinite impedance detector'

))) Here's what is sounds like - click to play the audio file (((

Here is the circuit diagram:

This detector has been a real eye opener for me in terms of its excellent performance, especially considering its circuit simplicity. Indeed in the past I have designed other more complex FET based infinite impedance circuits that do not quite work as well in practical terms as this latest circuit, at least according to my well calibrated ears!

I do not have access to any precise test equipment but my well calibrated ears tell me this 'two FET infinite impedance detector' is a beauty, surpassing practically every other AM detector I've built even at low RF input, and that's rather a impressive claim and the audio quality when used as an AM tuner feeding a high quality audio system is quite remarkable. Possibly the best thing about this detector is its excellent performance under weak signal conditions. Diode based detectors also work beautifully, but the use of an RF stage to ensure detection over a linear portion of the diode's curve is mandatory! This compound infinite impedance detector works

beautifully on the sniff of a useable RF signal.

Just add a high Q tuned circuit and that's it!

Felix Scerri
VK4FUQ
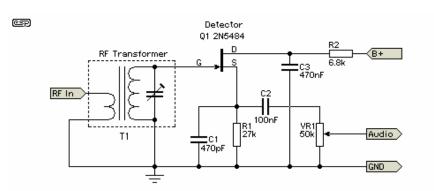
A better FET for the 'basic' Infinite Impedance Detector:

Quite recently by accident, I've realised the MPF102 FET that I've long used in my FET based infinite impedance detectors is possibly not the best FET to use. This was the reason why I developed the 'two FET' infinite impedance detector some time ago which works very well.

However I've found the choice of a more suitable FET works beautifully in the basic FET based infinite impedance detector circuit, which has appeared for many years in many editions of the ARRL Handbook.

I use the 2N5457 and others of the same 'family' may be equally suitable, but I haven't tried them! However with a 2N5457 in place of an MPF102, the basic infinite impedance detector has became my AM detector of choice. It works beautifully even at low signal input with lovely and clean low distortion audio along with a very high input impedance for good tuning selectivity. It's a beauty! The basic generic circuit is attached, courtesy of Rod Elliott's ESP website.

73 Felix VK4FUQ
10 / 02 / 2012.

The basic generic circuit is attached, courtesy of Rod Elliott's ESP website
Felix Scerri VK4FUQ

As often happens with me, my renewed interest in FET based 'infinite impedance detectors' of late has led to some interesting new research and I may have considerably improved the 'two FET infinite impedance detector' as a result.

My research suggests that although the use of a CCS (constant current source) reduces audio distortion in an audio stage, the value of the 'source resistor' in the CCS stage is somewhat critical for best results.

By using a potentiometer in lieu of a fixed resistor I have found that a resistance value of around 470 kohms cleaned up all overall audio distortion. I used an MPF102 as the CCS in this circuit. An interesting and worthwhile little circuit refinement.
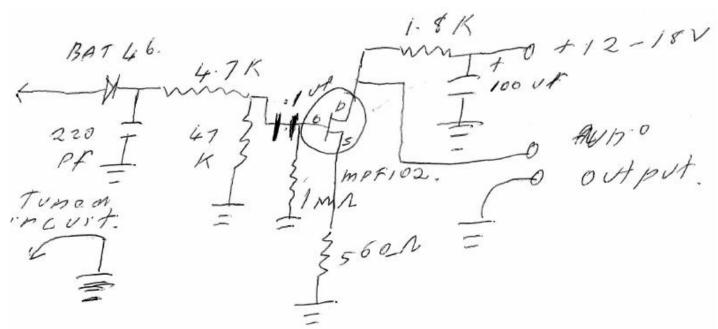
73 Felix VK4FUQ
21 / 02 / 2012.

A Minimum Component Count High Quality AM Detector
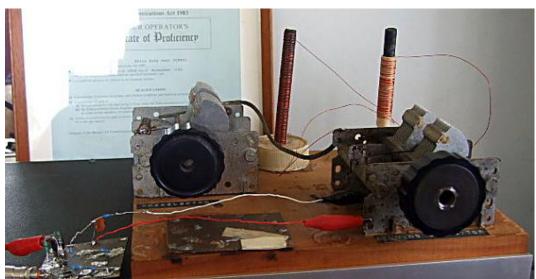
I was generally messing around with various circuit ideas and I came up with this AM detector circuit, a simple diode detector along with a FET stage. It was an attempt to provide good performance along with minimum number of components. Actually I've been pleasantly surprised at the excellent level of general performance and the best of all, it sounds great!

The circuit is quite conventional being a BAT 46 diode detector feeding an MPF102 FET buffer/ common source amplifier stage. I would ordinarily use a FET source buffer stage in this application, but opted to use a simple low gain FET 'common source' amplifier stage instead, with excellent results. I also used a BAT 46 Schottky diode instead of an ordinary germanium signal diode.
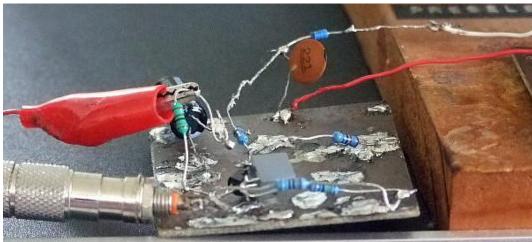
This was done for several reasons. Firstly, germanium diodes are now very hard to find but in any case these 'germanium diode equivalent' Schottky diodes are actually a superior diode, having very low noise, almost zero back leakage and an essentially complete absence of carrier storage effects and very good weak signal sensitivity. I call these diodes high fidelity diodes as they sound wonderful as RF detectors.

Circuit Diagram of the Minimum Component Count AM Detector by Felix Scerri

Minimum Component Count AM Detector by Felix Scerri

A Closer View

The high impedance of the FET's gate circuit is perfect for optimal buffering of the diode detector, something very important for good low distortion diode detector performance. Apart from providing slight voltage gain, the use of the common source FET amplifier is a new idea, as this prevents the possibility of incidental RF rectification occurring in a FET source follower stage, which can happen. A 1 uf plastic film capacitor may be added in series with the audio 'hot' output lead to block the DC offset out of the FET drain, if required.

Despite no additional RF stage ahead of the diode, audio quality on even relatively weak RF strength stations is actually very good, and of course the audio quality will be even better with increasing RF signal strength, something which will also increase the audio output level. Just on this, for a long time I was somewhat negative regarding diode detectors, as one AM station locally (the strongest one) was always distorted when using a diode detector.
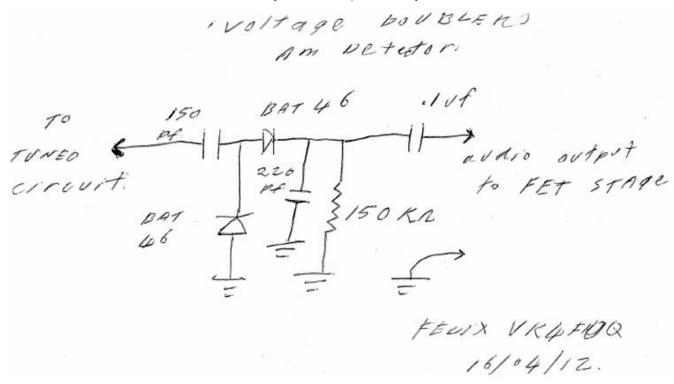
I strongly suspected a transmitter fault, but my complaints were ignored, until one day some time ago when all audio distortion suddenly disappeared! Nothing was ever 'said', but I realised that my suspicions of a long standing transmitter fault were correct, after all!

73 Felix
VK4FUQ
02 March 2012

Voltage Doubler Detector

This is an AM detector circuit that I've long known about and which worked ok, but never seemed to work as well as it should have. However I spent some time late last night trying to optimise the circuit, with some success.

It is a curious circuit being essentially a 'voltage doubler' originally developed for power supply applications, and its use as an AM detector is hard to analyse! It seems that the component values in the circuit are somewhat critical for good performance and if not, the performance is rather 'ordinary'. The circuit that I eventually came up with uses a 150 picofarad 'input capacitor' with a 150 kohm 'load' resistor and loaded into a FET common source voltage amplifier stage (as previously described) through a coupling capacitor with a 1 Mohm input resistance.

With these circuit values, it all works 'quite well'. Give it a go! It's an interesting AM detector with quite good 'sensitivity' and clean audio quality, and it seems to work well at low signal levels.

73 Felix
vk4fuq.

16th April 2012.

Simple AM detectors: What works best? A practical experimenters viewpoint.

I have written a lot about simple AM detectors for use as tuners for feeding into an audio amplifier, and it has been a long time interest. These days I use either diode based or 'infinite impedance' types of AM detectors. In this location our 'local' AM stations are quite distant and are therefore quite weak in terms of signal strength.

As such I find infinite impedance detectors based on field effect transistors give consistently better results for tuner applications due to their lower apparent overall detector distortion. Diode based detectors are quite 'fussy' as they require both optimal output buffering (AC/DC ratio) and an 'adequate' (beyond the diode knee) level of RF signal injection. http://www.tonnesoftware.com/appnotes/demodulator/diodedemod.html

Diode based detectors will happily 'detect' at very low signal levels, however the (inevitable) audio distortion that results, can be extremely irritating to the ear! Under these conditions I find infinite impedance detectors (even without additional RF preamplification and subject to individual FET characteristics), generally sound 'cleaner' and more pleasant to the ear.

FET's of course require a power source for operation whereas diodes are passive (un-powered) detectors (most of the time), however this is of no real advantage in a tuner application as an 'active' audio amplifier stage will generally be required anyway for audio level boosting, buffering etc.

In the end it will come down to a consideration of prevailing RF signal levels and other related circuit considerations at one's location. If local RF levels are strong, a well designed diode detector will give excellent results. If not, an 'infinite impedance' type of detector is most likely the better option unless one goes towards the option of additional RF preamplification prior to the diode detector.

73 Felix
vk4fuq.
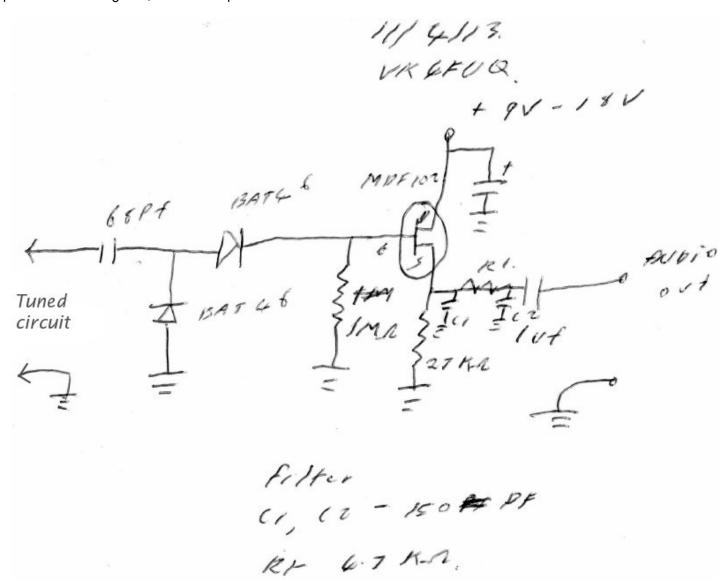
17th April 2012.

ANOTHER DETECTOR by Felix Scerri !

I've tried diode based 'voltage doubler' (or more correctly 'diode integrator') AM detectors before with indifferent results, however the other day, just trying a few ideas I came up with this version that works rather well, with low audio distortion, high audio output and really 'nice' audio quality and the best of all, it seems to work very at very low RF input level.

The two diode 'voltage doubler' detector using two BAT 46 silicon schottky diodes feed directly into a MPF102 source follower stage set at 1 Megohm input resistance. The 'input' capacitor feeding the diodes from a tuned circuit is 68 picofarads.

I have the simple RF filter right on the output of the FET stage. In that respect this circuit is vaguely similar to the old 'Selstead-Smith' valve AM detector of the past. An interesting one! I am very happy with its general performance. Regards, Felix vk4fuq 11/04/2013

Felix Scerri VK4FUQ

UPDATE - JUNE 2013

G'day all, readers may recall the two FET infinite impedance detector I developed some time ago. That circuit worked well, but some samples of the MPF102 regretfully produced distorted output. However a recent discovery has resulted in an improved version that has truly exemplary performance.
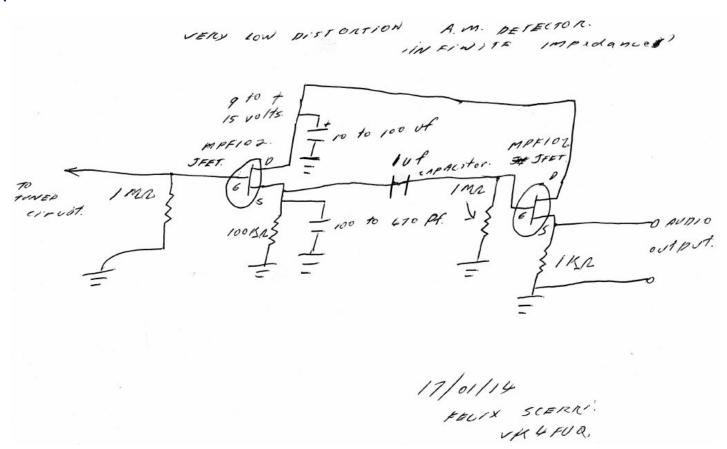
The MPF102 is in all honesty a device essentially designed for VHF applications, but is often used in simple audio applications albeit with occasionally indifferent results, due to general device parameter 'spread'. More or less by empirical trial and error I have found that a simple change in source resistor value to 100 k (from a much lower value), pretty well fixes everything.

In a FET source follower circuit this is interesting as the output impedance is actually a lot lower than 100 k due to the action of the FET's transconductance. It is similar to the action of a bipolar transistor emitter follower circuit. In fact a simple infinite impedance detector with a 100 k source resistor actually works very well and will be good enough for many applications, however the addition of a second FET as a constant current source does markedly reduce overall audio distortion in the stage overall, and produces very clean and low distortion audio quality and also reduces the output impedance considerably (good for tuner applications), so take your pick!

Regards, Felix vk4fuq 02/06/2013.

UPDATE - JANUARY 2014
A simple modification to the basic FET 'infinite impedance' AM detector that dramatically improves performance.

G'day all, over the Christmas break just messing around I worked out (mostly by accident), a simple modification for the simple FET based 'infinite impedance' detector circuit that dramatically improves weak signal performance and also greatly reduces audio distortion.

Essentially by the addition of another FET 'buffer' stage, another source follower, capacitively coupled from the first detector stage. The circuit is actually a simplified version of the circuit that I described in this link, http://sound.westhost.com/articles/am-radio.htm (figure 6) and testing the two head to head, they both sound superb and the simpler version is actually somewhat easier to build. I cannot get over how low distortion and 'nice' the recovered audio sounds. It is a joy to listen to!

Regards, Felix vk4fuq 12/01/2014.

Diodes for 'weak signal' crystal set applications.

As I am primarily interested in using crystal sets as AM tuners for feeding into a preamplifier/amplifier and loudspeakers, the actual type of diode can be relatively critical. In a strong signal area, not so much, but in a weak signal area such as where I reside, definitely. A diode with a good 'square law' performance (the area below the 'knee' in the diode curve), generally results in much cleaner and lower distortion than other good performing diodes, and believe me diode distortion under weak conditions, even with optimised diode 'buffering' used is very nasty sounding to the ear!

Testing many, many diodes in actual working crystal sets and listening critically to the audio output, it seems to me that the best diodes to use under weak signal conditions are the so called 'gold bonded' germanium diodes. I have sampled many different gold bonded germanium diodes and they all work well in this specific application, although sometimes the rectified output voltage may not be as high as other germanium or silicon schottky diodes, however the audio quality is much cleaner and shows much less apparent distortion!

I have tried OA5, OA47, IN141 and several 'CG' gold bonded germanium diodes with consistently excellent results. Other 'ordinary' germanium diodes may also work well under weak signal conditions, but they will need to be tested individually to check actual performance in a working circuit. One thing that I have noticed about germanium diodes is that due the 'point contact' nature of their construction, even diodes of the same specific type can exhibit rather different levels of performance!

The otherwise very good BAT46 silicon schottky diode works extremely well at good RF input level but not so well at weak RF input, especially when heavy broadcast 'processing' (commonplace these days), is used. New 'gold bonded' germanium diodes are probably no longer made although I am aware that they can be purchased through vendors over the internet. Apart from that, they may be found in old gear. As stated earlier, ordinary germanium diodes may be quite good but will need to be checked individually. Diodes are complex things!

Felix (vk4fuq). 29/01/2014.

* * * * * * * * * * * * * *
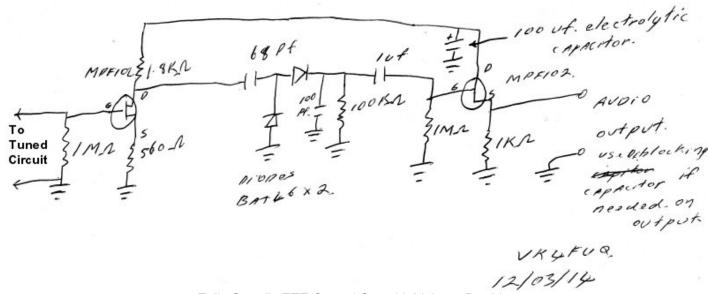
Just when I thought that I've tried everything I realised that I've not tried a voltage doubler (diode integrator) detector with an RF stage in front. It is an unfortunate fact that all diode detectors 'need' a good RF injection level (and buffering) or audio distortion becomes 'bad' (an understatement!)

Consider yourself very lucky if you live in a strong signal area! (unfortunately I don't). Anyway I quickly built a couple of prototypes which worked ok but not as well as hoped. Initially I tried using a source follower buffer (no voltage gain), but converting it to a common source amplifier (with appreciable voltage gain), seemed no better, which was strange.

After staring at my prototype for what seemed like forever with my magnifiers on, I (finally) realised my mistake. A common source stage takes its output off the 'drain' terminal of the FET, not the 'source' terminal as it does with the FET source follower buffer stage. I shifted one wire, and all worked as expected and it sounds fantastic! All diode detectors (one diode envelope detectors, doubler detectors etc), need good RF input and a simple untuned FET RF gain stage works very well. As to the sound, it sounds great.

73 vk4fuq 10/03/2014.

Felix Scerri's FET Crystal Set with Voltage Doubler

A 'nice sounding' AM detector - update October 2015

Looking through these pages the other day I was slightly shocked to realize that I've completely forgotten about some of the circuits that I've contributed in earlier days! They all work pretty well, however in my advancing older age and a little like my tastes in high fidelity generally, I'm starting to show a particular preference for 'nice sounding' bits of audio gear.

I guess that this also means low noise/low distortion too, although with AM broadcasting, at least in this country and probably elsewhere in the world too, the very common use of broadcasting 'processing' tends to make it hard for AM detectors generally, often resulting in a 'hard/compressed' sort of sound although still low in distortion, is not 'nice sounding', if that makes any sense!

Well of all the AM detectors that I've tried and/or developed, only one sounds 'nice' when confronted by heavy broadcast processing and that is an AM detector 'based' on a voltage doubler/diode integrator circuit (similar to the above circuit, actually the 68 pf capacitor should be changed to .1 uf, for slightly greater output).

It has taken me a very long time (years) to 'optimise' this circuit, but as it presently stands this is my favourite 'nice sounding' AM detector, and gets most use for general high quality listening on the AM broadcast band. It sounds really good! It somewhat reminds me of an old OA47 gold bonded diode detector from years ago before such abhorrent 'processing' became commonplace!

Ah yes the OA47….now that is a lovely sounding detector diode!

Regards, Felix vk4fuq. 17 / 10 / 2015.

* * * * * * * * * * * * *

That's it for crystal sets. I hope you try building one, it's easy and great fun!

See some useful links below....

73's
Mike

Crystal Sets (Part1) | **Build Your Own Crystal Set (Part 2)**

Spider's Web Crystal Set (Part 3) | **Crystal Set By Kenneth Rankin (Part 4)** | **Crystal Radio Links**

No AM radio stations or transmitters in your locality or country?

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000 Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build:
http://www.sstran.com/pages/products.html

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo Complete, high quality ready built medium wave AM Transmitters from Vintage Components: http://www.vcomp.co.uk/index.htm Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

METZO AM Medium Wave Transmitter with built in compressor:

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS: Fine links to more Crystal Radio websites here

Component Suppliers: Links to electronic component suppliers here

Visit my Amateur Radio Pages >

Crystal Sets (Part1) | Build Your Own Crystal Set (Part 2)

Spider's Web Crystal Set (Part 3) | Crystal Set By Kenneth Rankin (Part 4) | Crystal Radio Links

^Top Of Page

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables

www.mds975.co.uk © 2004 - 2015

CRYSTAL SETS 2
Some Practical Designs
MAKE YOUR OWN
CRYSTAL SET !!

Home | **Contact** | **Site Map** | **Radio Stations & Memorabilia** | **Amateur Radio**

Crystal Sets Introduction | **Resistor & Capacitor Conversion Tables**

Crystal Sets (Part1) | **Spider's Web Crystal Set (Part 3**

Crystal Set by Kenneth Rankin (Part 4) | **Experimental Crystal Sets (Part 5)** | **Crystal Radio Links**

CRYSTAL SETS 2: SOME PRACTICAL DESIGNS
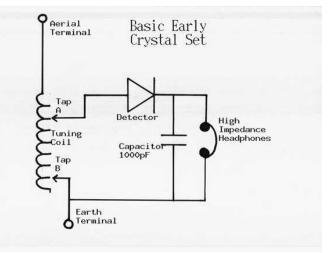
I hope that you attempt building one or two of these crystal set designs and I really do recommend that the components are carefully connected up using soldered joints onto a piece of tag-strip for reliability. However if you are new to constructing such electronic circuits then some simple solder-less techniques could be employed and these are suggested at the bottom of the page.   Also see Crystal Sets Part 5 for more ideas on experimenting with crystal sets.

An early and very basic crystal set would have been nothing more than a coil of wire, perhaps 50 -100 turns, wound around a cardboard tube about 3 inches (7cm) in diameter, a detector (or cats whisker) and a pair of special High Impedance headphones (as discussed in part 1).

There would be a very large aerial strung up around the garden and the all important connection to earth.

The coil would have tapping points (connection points) at intervals of around 5 or 10 turns. See the circuit diagram on the right for details of who the set is wired together.

The tapping points on the coil allow the set to be tuned to different frequencies by adjusting the position of tap B. Tap B would be connected to the coil at differently positions by way of a crocodile clip. The fewer turns between the top (aerial end) of the coil and tap B, the shorter the wavelength received (ie the higher the frequency). Tap A would allow the detector to be connected at different positions to vary performance.
There is an additional component drawn in the above diagram, the *capacitor* (value 1000pF), this is included in crystal sets that used the *High Impedance* magnetic headphones, and bypassed any remaining radio frequencies (RF) to earth.
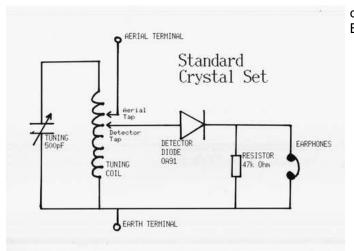
A very basic crystal set circuit.

I have not built the set described above as it is so basic. Such a crystal set above would probably have been adequate in 1920 - 1923 when there would have been only one local transmitter receivable.

When the BBC expanded transmissions and it became possible to hear more than a single station it would have became necessary to include a more convenient means of tuning the set.

This was achieved by including a Variable Tuning Capacitor, of about 500pF (0.0005uF) connected in parallel with the tuning

coil forming a *tuned circuit*. The tuning capacitor would have a Bakelite knob on the spindle to aid tuning.

The Standard Crystal Set

Because of the simplicity of crystal sets, it is often difficult to separate stations. When tuned into one station it is often possible to hear another close by station in the background, this is due to lack of *selectivity*. This can be reduced somewhat by adjusting the positions of the Aerial Tap and Detector Tap. Moving them closer to the bottom of the coil, the earthy end, reduces the load on the tuned circuit and this improves selectivity, however it does also reduce *sensitivity* which can make the station quieter.   Headphones will often swamp a tuned circuit and reduce its selectivity (Q factor), so moving the tapping point lower down improves this situation. Every circumstance is bound to be different though so the best balance has to be found by experimentation. My crystal set has both the diode and the aerial connected to the same tapping point on the coil, about a quarter of the way down.

The modern 'standard crystal set' shown above uses a Crystal Earphone, since suitable high impedance magnetic headphones (of 2000 to 4000 ohms) are no longer widely available. When using a crystal earpiece the 1000pF capacitor shown in the first diagram can usually be omitted an in its place a 47k ohm resistor is connected, this ensures that the Crystal Earphone will work at its most efficient i.e. the sounds will be as loud as possible. The resistor allows DC current to flow through the circuit efficiently - this would otherwise be blocked when using a crystal earphone.   In a modern crystal set the detector used is a *Diode.* Suitable diodes include OA80, OA81, OA90 OA91 and IN94 which are usually available from component stockists.

A Better Diode For Increased Efficiency
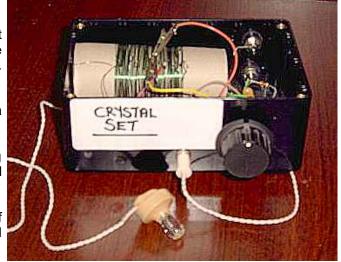
The OA47 will be of particular interest since it has the lowest forward bias voltage of any of these diodes which will make the crystal set somewhat more sensitive and therefore louder. The US equivalent of the British OA47 is the IN34.

On the right you will see my real working example of a crystal set

The large plastic knob on the front turns the variable tuning capacitor. This set receives the three UK national stations and also three local radio stations very well at my location.

There is a small 3.5mm jack socket mounted on the front of the plastic case (MB5 from Maplin Electronics) that the crystal earphone plugs into.

The coil can be seen inside the case, it is 70 turns of 30 gauge enamelled copper wire wound around the centre of a toilet roll and tapped every 10 turns, by scraping off the enamel insulation and making a small twist. The croc' clips can be seen clipped on to these twists to connect to the aerial and detector tap points.

A real working crystal set. Radio as if by magic with no battery or mains power.

THE MEDIUM WAVE COIL - MORE DETAILS
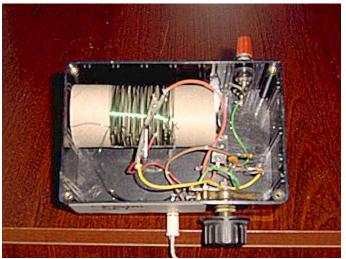
**PHOTO SHOWING THE INSIDE OF
THE COMPLETED CRYSTAL SET**

Medium Wave Coil

The number of turns of wire required on the coil will vary depending on the size of the former (in this case the inside toilet roll) and the thickness of the wire. So to obtain the correct coverage of the medium wave band may need a little experimentation.

I usually find that between 50 to 90 turns is right and I generally use enamelled copper wire that is between 30 s.w.g. and 26 s.w.g (i.e. 0.315mm and 0.45mm diameter), so it's best to start with too many turns and then work down.

The more turns that you use the lower the frequency range will be, i.e. too many and the coverage of the top end of medium wave around 1500 - 1600 kHz will be lost, while too few and the coverage down to 500 kHz will be lost.

It is also important that the coil former is non conducting, i.e. not metallic. It could be wood or cardboard or a short piece of PVC piping and with a diameter of between 1½ and 4 inches (4 to 15 cm) are common sizes. You could try using a ferrite rod too, see below.

This particular set has a coil wound onto a toilet roll tube which consists of 70 turns of 30 s.w.g. (0.315mm dia) enamelled copper wire tapped at every 10 turns. It also has the additional small trimmer capacitor that helps match the aerial to the tuned circuit thereby improving selectivity, see below.

USING A FERRITE ROD AS THE COIL FORMER

The aerial coil could be wound onto a ferrite rod.

A piece of 10mm diameter ferrite rod of between 3 and 6 inches long (80 to 150mm) will be most suitable and will require between 50 and 90 turns of enamelled copper wire to provide coverage of the medium wave band: First make a paper tube that is held together with sticky tape that will easily slide up and down the ferrite rod. Then wind the coil over this with the windings neatly side by side. Make tapping points every 10 or 15 turns so that the aerial and diode tapping points can be adjusted.

Adjustments to the tuning range can be made by removing some wire from the coil so it is best to start off with too many turns and then work down. Fine adjustments can be made to the completed coil by sliding it up and down the ferrite rod.
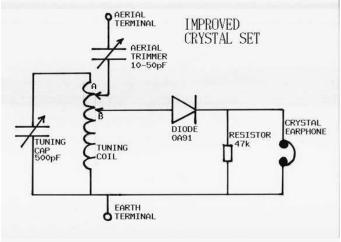
AN IMPROVEMENT TO THE DESIGN

The crystal set above also has one small, but significant, improvement over the standard crystal set and that is an Aerial Trimmer. A trimmer is a variable capacitor, very similar to the tuning capacitor, except smaller and adjusted with a screwdriver.
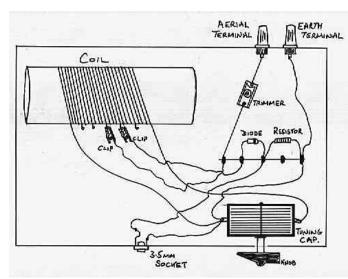
The value of the trimmer is usually around 10 - 50pF, but if a small tuning capacitor is available that will probably be just as effective. In the absence of such a variable capacitor, individual fixed ceramic capacitors of e.g. 10pF, 50pF and 100pF can be tried in this position to judge which gives the best results with the particular aerial being used.

The trimmer capacitor adjusts the coupling to the tuned circuit, reducing the load of the aerial on the tuned circuit will improve

the selectivity (Q), and it will be easier to separate stations. Again tapping points are used and I find this to be an excellent arrangement.

Improved Crystal Set design, with good selectivity

Layout Of The Crystal Set - **Although this is soldered together an alternative to tagstrip would be a 5amp mains connector block so that components can be trapped in place with screws. See article below.**

The picture on the right shows the general layout of the crystal set above. The coil is of approximately 70 turns is wound on the centre of a toilet roll, and has tapping points at 10 turn intervals.

The trimmer is soldered between the Aerial terminal and the piece of 5-way tag strip, and a wire goes from there to a croc' clip which is clipped onto a tap on the coil. The Diode is also soldered onto the tag strip, one end connected to a piece of wire going to a second croc' clip & connected to a tapping point on the coil, the other end of the diode is connected to the 3.5mm jack socket that the Crystal Earphone plugs into.

The 47k resistor is also connected to the earphone end of the diode and goes to earth, the earth terminal wire is soldered to the tag strip at this point too. The tuning capacitor has two terminals, one connected to each end of the coil, and one of them is also connected to earth as shown. [Where the wires cross over in the diagram, they do not touch and are not connected together].

LONG WAVES

In most areas around Europe and certainly around much of the UK you will be able to hear a Long Wave station. To receive Long Wave on a crystal set will require an aerial coil with a greater number of turns to increase its inductance.

As a good general guide a coil wound on a piece of 10mm diameter ferrite rod will require about 250 turns of enamelled copper wire: First make a paper tube that is held together with sticky tape that will slide up and down the ferrite rod. Then wind the 250 turn coil over this, the windings will have to be made over the top of each other. Make tapping points at, say, 50, 75 and 100 turns to tap the aerial and diode to.

As with the medium wave ferrite rod aerial, adjustments to the tuning range can be made by adding or removing some wire from the coil, and fine adjustments can be made to the completed coil by sliding it up and down the ferrite rod. The longer the ferrite rod the better and anything between 3 and 6 inches long (80 to 150mm) will be very good.

SHORT WAVES
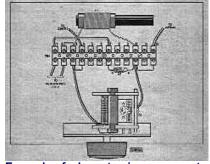
If you like experimenting, then reducing the number of turns on the coil to say 10 to 30 will allow reception of the higher frequencies, the Short Waves. I have found

that winding the coil around a 'ferrite rod' often works even better with short wave reception.

Obtain a ferrite rod about 7 to 15 cm long and about 1cm in diameter. Make a couple of small tubes of card, about 4cm long, that will fit tightly over the rod.

On one tube wind two coils using 0.5mm diameter enamelled copper wire - one coil of about 30 turns and a second one of 2 or 3 turns wound over the top of the first. Secure the windings in place with Sellotape.

On the other card tube wind a similar coil, but use about 15 turns for the first coil and for the second coil wind about 3 to 4 turns over the top, and secure with Sellotape tape.

Example of a layout using a connector block to wire up a crystal set
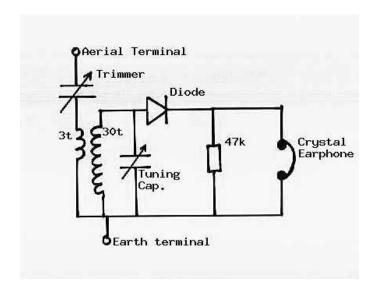
These coils will provide coverage of short wave in two bands using the first coil for the longer wavelengths, typically 60 to 31 metre bands and the second coil for the shorter wavelengths typically 25 to 19 metre band. Wire up the circuit as shown in the circuit diagram below.
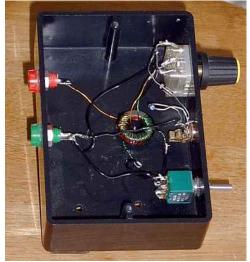
USING A TOROID INDUCTOR FOR SHORT WAVES

Even better selectivity performance can be achieved by winding the inductors (coils) on a ferrite toroids (T50-2 yellow, or green will do). The aerial trimmer need not be used if selectivity and sensitivity is found to be adequate. It's all about experimenting, and I find it best to use a trimmer or small coupling capacitor to obtain the best selectivity.

Up to 30 turns of 0.5mm enamelled copper wire can be used for the longer short waves below 10 MHz, while a winding of around 15 turns will provide coverage of the shorter short waves above 10MHz.

The circuit diagram of the Short Wave Crystal Set

A completed SW Crystal Set using a toroid inductor. Note: the main winding has a tap to allow the switch to short part of the winding and thereby give two ranges.

AUSTRALIAN DESIGN

Moving back to the Medium Waves, here is a circuit for a very interesting Australian design that promises extremely good station separation (selectivity), and having built it I can vouch for that claim, it's really excellent.

I receive three national stations and three local stations at my location with excellent clarity using a modest antenna and standard crystal earphone.

The coil is different to the other crystal sets described above, it is
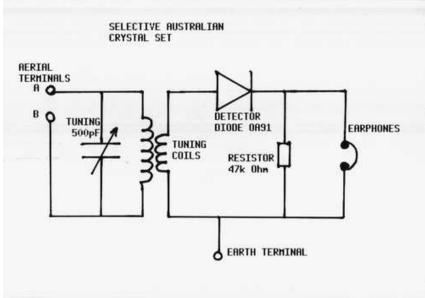
much bigger at 3¼ inches (8cm) diameter and 5inches (12cm) long. I made my coil former out of the cardboard from a breakfast cereal box - just like Blue Peter!

The design is often referred to as The Mystery Crystal Set, by Proton.

The front panel of the Australian Crystal Set

Two distinct coils are wound on it, the first one consists of about 50 turns of 24 s.w.g (approx) enamelled copper wire. The second coil is 25 turns, very close wound right over the top of the first coil using 30 s.w.g. (approx) wire, try to get this second coil wound in between the windings of the first, for better inductive coupling.

Then carefully wire up the set according to the diagram. Notice that the tuned circuit is not connected to earth and has no *direct* connection to the detector circuit. The detector circuit *is* connected to earth however. The two aerial terminals offer alternative selectivity performance, terminal A gives very good selectivity while B is very wide. I never bother with B.
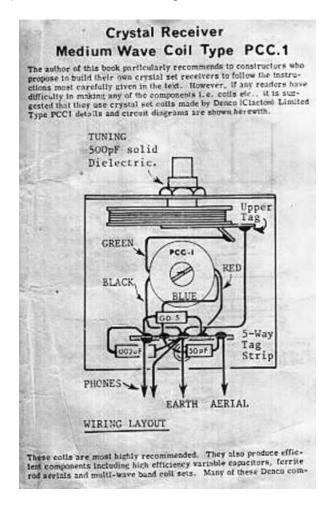
Make the coil carefully and wire up this crystal set according to the circuit diagram opposite and you will be rewarded with a really high performance crystal set of a type that was used in the very early days of broadcasting in 1930's in Australia.
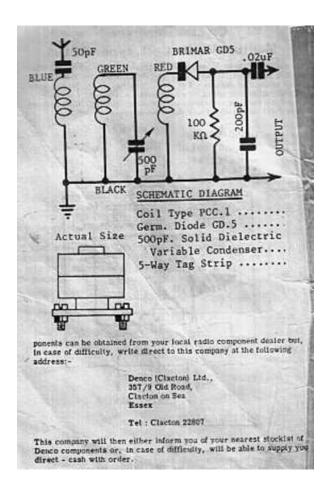
This is probably my favourite crystal set!

THE DENCO PCC1 COIL

The PPC1 coil was a commercially manufactured by Denco Clacton Ltd and was popular among hobbyists not keen on going to the bother of winding their own fiddly little coils. As a child I wanted try one of these coils and sent away for one by mail order. It arrived a few days later in a little cloth bag, like a miniature pump bag, with protective wrapping inside.

The coil windings are entirely enclosed in what I can only describe as a cylindrical ferrite 'shell', the four very thin connecting wires exiting, two either side, from small apertures in the 'shell'. The performance of the circuit shown below I seem to remember was quite pleasing. Unfortunately I cannot find the set or the PPC1 coil at the moment, but here is

a reproduction of the circuit diagram and data:
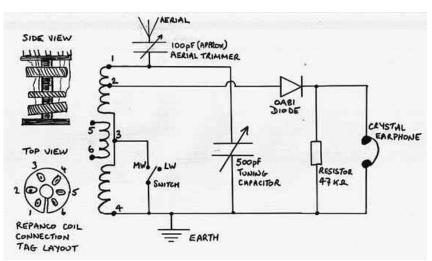
THE REPANCO DRR2 COIL

I recently rediscovered an old Repanco DRR2 Longwave / Mediumwave coil that must have been kicking around in my junk box since the 1970's.

The DRR2 coil was made by Repanco in Coventry. It came with a page of suggested circuit diagrams which I thought had been lost to the mists of time, but it recently came to light again, so I have now copied it below.

Once again I included an aerial trimmer which can be adjusted to improve selectivity.

The circuit diagram of the crystal set using the Repanco DRR2 coil

Repanco Ltd was formed by two ex-army signals engineers and from the earliest days of radio supplied crystal set kits and coils to radio construction enthusiasts.

The Repanco DRR2 coil was for medium wave and long wave intended for use in when building simple crystal set and valve radio
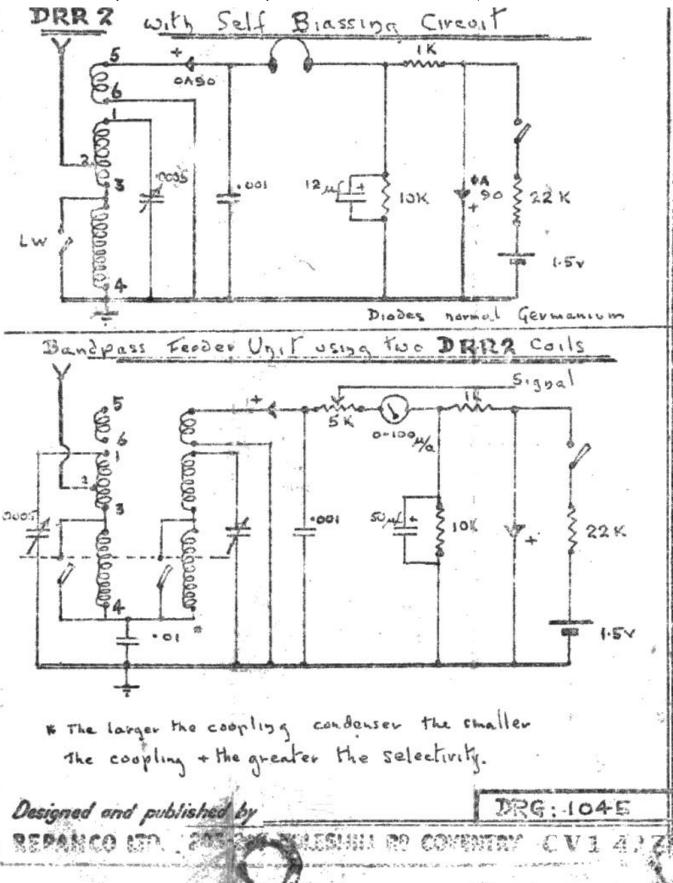
A 'lash-up' of the Repanco crystal set

circuits.
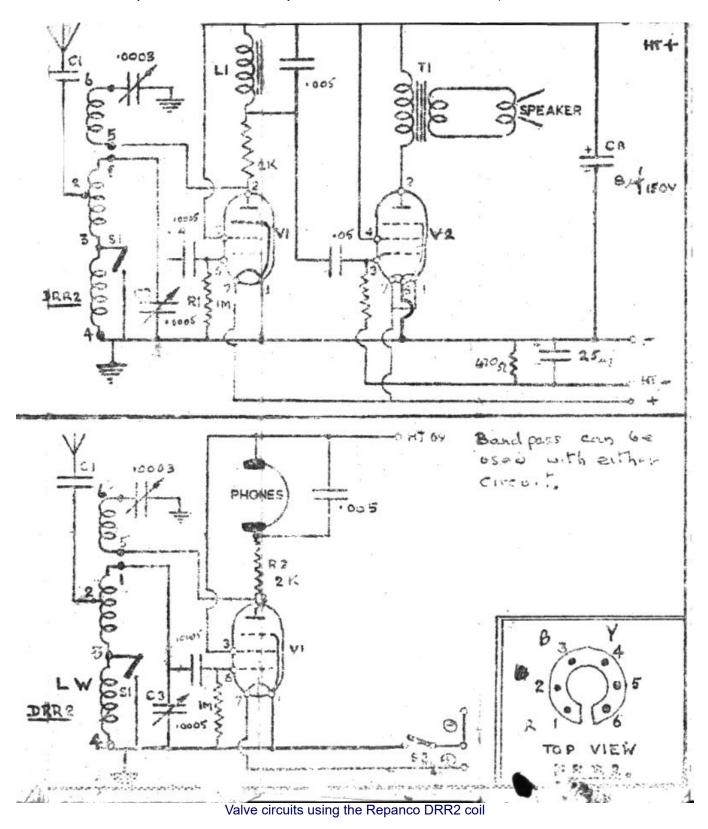
It consists of three coils; a Medium Wave coil at the top that includes a tapping point (for the aerial); a coupling coil or tickler in the middle; a lower coil which can be connected in series with to top coil to provide Long Wave reception.

I have built a quick crystal set with the coil and it provides good reception with excellent selectivity, so it must have a very good Q factor.

The Repanco DDR2 coil was provided with a simple Foolscap size information sheet that showed four different radio circuits. Sadly the sheet does not give a huge amount of information and my copy is now rather tatty and faded - it is copied below:

Simple crystal radio type circuits using the Repanco DRR2 coil

Valve circuits using the Repanco DRR2 coil
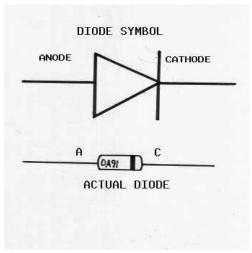
Repanco Ltd no longer produces radio coils and crystal set kits for the radio construction enthusiast, as it did in the early days of radio. In 1986 it was renamed Repanco Bartlett Ltd when it merged with Bartlett Electronics. The company moved from the Foleshill Road to Unit 24, Albion Industrial Estate, Endemere Road, Coventry CV6 5NT and now specialises in transformers and wound components and can design and manufacture to commercial customer requirements, their website is: http://www.repancobartlett.co.uk/
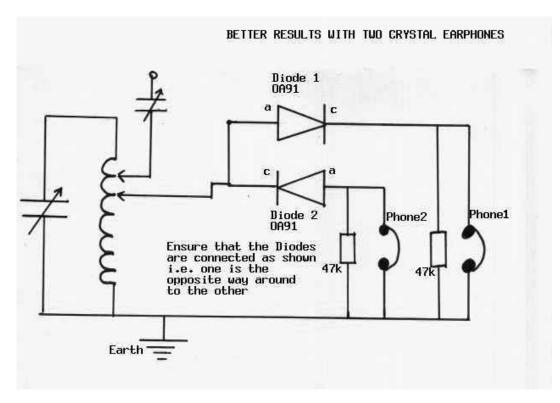
CRYSTAL EARPHONES

Here is a good idea and well worth trying, to maximise the use of sound output from your crystal set why not use dual crystal earphones? Having an earphone in each ear helps to block out extraneous noises helping the listener to better concentrate on any weaker stations received.

Using the circuit below, one earphone makes use of one half cycle of the radio wave while the second earphone uses the other half cycle of the wave that would have previously gone to waste when using just one diode. Ensure that the diodes are connected up according to the diagram i.e. one diode is connected the opposite way round to the other.

Also try to make sure that the diodes and crystal earphones are similar to obtain the best results. (You could simply connect two crystal earphones to the same terminals of the single diode, but this would not be as efficient and the sounds would be much quieter.)
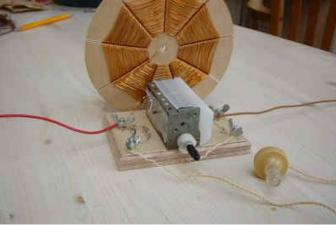
A note about Crystal Earphones: It will be worthwhile buying several different ones from different sources as performance varies between manufacturers quite markedly. I have found the ones marked 'Japan' on the back are the most sensitive and therefore loudest, whereas ones marked 'Receiver' 'Taiwan' are often a little less sensitive and therefore quieter and sometimes more 'tinny' sounding.

As mentioned previously it has been noted that the OA47 diode will be of particular interest since it has the lowest forward bias voltage of any of the common diodes available. This will make the crystal set somewhat more sensitive and therefore louder. The US equivalent of the British OA47 is the IN34.
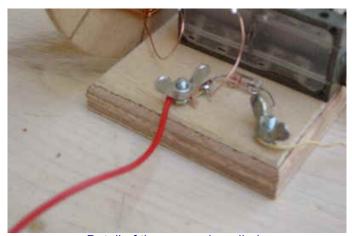
SPIDERS WEB' COIL

Here is an interesting concept sent in by Chris Dorna of the Vught North Scouts in the Netherlands. It is a crystal set made out of a coil wound in the form of a spiders web:
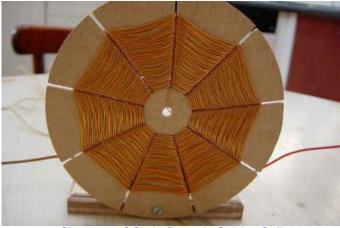
See more HERE.

Chris Dorna's Crystal Set with Spider Web Coil

Detail of the germanium diode

Close up of Chris Dorna's Spider Coil

LOOP THE LOOP!

A crystal set can also be made that does not need a large long wire aerial. If you have ever made a loop aerial for medium wave or long wave DX-ing, then it is a simple matter to add a diode, resistor and a socket to connect a crystal earphone that will allow reception of nearby stations.

See my section on Loop Aerials and ATU's for more constructional details.

A portable loop aerial, that incorporates a crystal set

The circuit diagram of the Loop Crystal Set. The loop is 10 turns of 7/0.2mm 'hook-up' wire wound on a 40cm (17") former made of attractive plastic edging strip available from many DIY stores. The loop is very directional in its pick up pattern, which can help eliminate interference from some stations by rotating the loop. The switch and additional capacitor allow tuning of the lower medium wave band from about 650 to 520 kHz. Having a loop with 50 to 60 turns of wire will tune into the Long Wave band.

DIODES - For Crystal Set Use - some notes by Felix Scerri

Germanium diodes for crystal set use.

Although I'm a fan of these new silicon schottky BAT 46 diodes, good germanium diodes still have a lot to offer, especially in terms of 'weak signal' sensitivity. Last night I did an experiment.

I sorted through quite a few of my hundreds of acquired random germanium diodes looking for particularly 'sensitive' ones. I tested this by tuning in a weak AM station and comparing the detected DC output level and also the apparent 'loudness' of the audio signal.

Even amongst germanium diodes of the same type, there was enormous variation all the way from excellent to poor! For very weak signals, germanium diodes 'detect' in the 'square law' region below the diode conduction 'knee', in a rather different part of the curve than with much stronger signals (way beyond the diode knee).

When testing germanium diodes for weak signal sensitivity, the inherent capacitance of the diodes is also a factor, and the 'tuning' may change somewhat and will need to be readjusted with every diode tested!
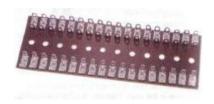
In the end, out of a large number of germanium diodes tested, I found three or four germanium diodes with excellent weak signal sensitivity and the rest were poor. One other interesting thing, good germanium diodes 'sound' different, rather more 'rounded and smoother' than the schottky's which tend to sound mercilessly clean, almost clinical. I also found almost no variation in weak signal sensitivity with my BAT 46 schottky diodes. Take your pick!

Regards, Felix Scerri VK4FUQ 14/03/2012

SOLDERLESS CONSTRUCTION IDEAS

For a novice the use of a soldering iron may seem a bit daunting at first and while the most reliable results will be obtained with a good soldered joint using a tag strip as shown below, the circuits can still be made without the use of a soldering iron.

5 WAY TAG STRIP
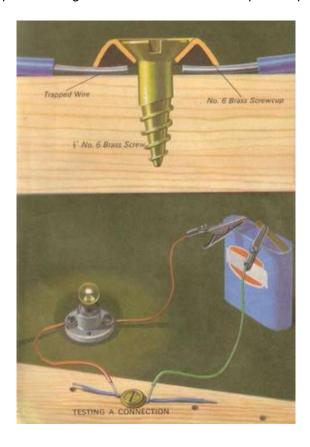
36 WAY TAG STRIP - TWO ROWS

The very simplest circuits could be wired together ,with a little ingenuity, with the component wires being held together in the grip of solderless crocodile clips, whereby the connecting hook-up wire is fixed to the croc' clip by a screw rather than solder.
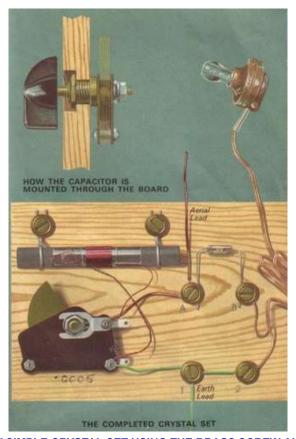
For more the slightly more complex circuits a plastic Terminal Block (sometimes referred to as a choc' or chocolate block) can be utilised very effectively indeed. These are used in mains wiring and are available in various sizes; 2 Amp, 5 Amp, 15 Amp and 30 Amp. The 5 and 15 Amp Terminal Blocks I have found to be the most suitable. The various component wires can be trapped securely with the screw at each junction point. This method also makes it easy to change the components around when experiment with different circuits. See The EXPERMENTAL CRYSTAL SET for more details in Part 5.

'CHOCOLATE' TERMINAL BLOCK

The Ladybird book called 'Making A Transistor Radio' (also shown on the TRF Radio pages) detailed a very novel approach using brass screws with screw-cups to trap the component wires at each junction point:

THE BRASS SCREW AND SCREW-CUP METHOD OF CONSTRUCTION

A VERY SIMPLE CRYSTAL SET USING THE BRASS SCREW AND CUP METHOD

Crystal Sets Part Part 3 >

No AM radio stations or transmitters in your locality or country?

Has your local medium wave broadcast station closed or been moved to VHF/FM or Digital? Don't worry. You can still build and experiment with crystal sets and TRF radios by also buying or even building a simple low power AM transmitter. So, not only can you use your crystal sets but you can also run your own radio station that can be heard in and around your home - playing the music or programmes that you want to hear!

SSTRAN AMT3000: Superb high fidelity medium wave AM transmitter kits from SSTRAN. Versions available for 10kHz spacing in the Americas (AMT3000 or AMT3000-SM) and 9kHz spacing in Europe and other areas (AMT3000-9 and AMT3000-9SM). Superb audio quality and a great and well designed little kit to build:
http://www.sstran.com/pages/products.html

http://www.sstran.com/

Other AM transmitters available:

Spitfire & Metzo: **Complete, high quality ready built medium wave AM Transmitters from Vintage Components:** **http://www.vcomp.co.uk/index.htm** **Vintage Components offer a choice of the high quality Spitfire and Metzo transmitters:**

SPITFIRE AM Medium Wave Transmitter with 100 milliwatt RF output power:

http://www.vcomp.co.uk/spitfire/spitfire.htm

METZO AM Medium Wave Transmitter with built in compressor:

http://www.vcomp.co.uk/metzo/metzo.htm

AM88 LP: A basic AM transmitter kit from North County Radio.
http://www.northcountryradio.com/Kitpages/am88.htm

LINKS:

BOWOOD ELECTRONICS - A friendly, helpful and very speedy source for many of your electronic components at prices that won't frighten your wallet!

THE FOXHOLE and P.O.W RADIOS - Simple crystal set receivers used by soldiers during the war and by prisoners of war (P.O.W.'s).

VINTAGE COMPONENTS - A great resource for crystal sets, components, valve radio kits and medium wave AM transmitters!

6V6 - Electronic Nostalgia and Vintage Components

Crystal Sets Part 3 >

Crystal Sets (Part1) | Spider's Web Crystal Set (Part 3

Crystal Set by Kenneth Rankin (Part 4) | Experimental Crystal Sets (Part 5) | Crystal Radio Links

^Top Of Page

Visit my Amateur Radio Pages >

Home | Contact | Site Map | Radio Stations & Memorabilia | Amateur Radio

Crystal Sets Introduction | Resistor & Capacitor Conversion Tables
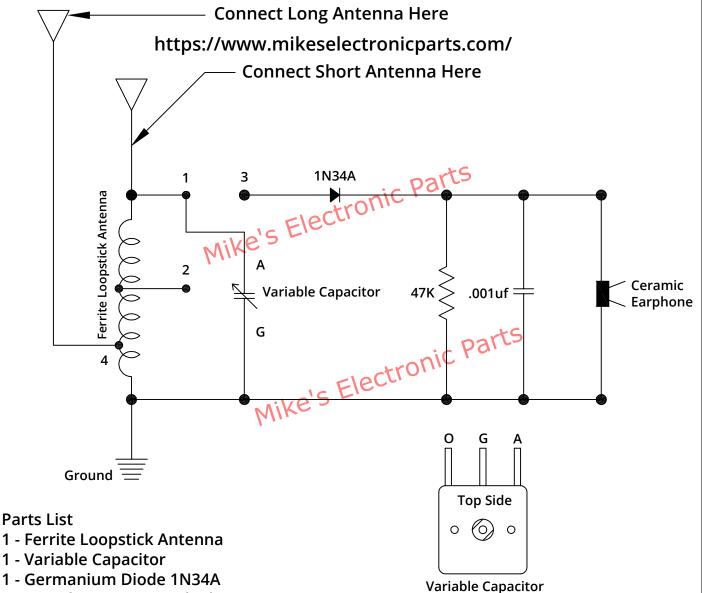
www.mds975.co.uk © 2004 - 2014

Mike's Electronic Parts, LLC
Simple Crystal Radio Kit 2

Connect Long Antenna Here

https://www.mikeselectronicparts.com/

Connect Short Antenna Here

Ferrite Loopstick Antenna

1N34A

Variable Capacitor

47K   .001uf

Ceramic Earphone

Ground

Mike's Electronic Parts

O   G   A

Top Side

Variable Capacitor

Parts List
1 - Ferrite Loopstick Antenna
1 - Variable Capacitor
1 - Germanium Diode 1N34A
1 - .001uf Capacitor Marked 102
1 - 47K Resistor
1 - 20 Million Ohm Ceramic Earphone
Antenna and Ground wire not include with parts

For short antenna connect a 15ft to 30ft length of wire to "1". For long antenna 40ft to 100ft connect to "4". It is best to use a long wire antenna. If you are not using the taps do not cut the wires. Doing so will cause the coil not to work. Various connections are, Connect 1 to 3 or Connect 2 to 3. Most volume can be had at 1 to 3 less volume but better selectivity connect 2 to 3 and the antenna connected at 1. It is best to solder the coil wire, not doing so may cause a bad connection and the radio will not work.
Drawn By: Scott Lowe
This Kit can be bought at my web site Part Number CRK#2 Kit.

https://www.mikeselectronicparts.com/

THE SMITHSONIAN INSTITUTION

The Smithsonian Institution is home to more than 141 million objects, ranging in size from insects and diamonds to locomotives and spacecraft. It is the world's largest museum complex, comprising 15 museums and galleries and the National Zoo in Washington DC, and two additional museums in New York City. Millions of visitors each year visit the nation's capital to view such treasures as the Hope Diamond, the Star Spangled Banner, and the Wright Flyer. A broad range of exhibits ensures a fun and educational experience for young and old alike.

One of the world's leading scientific research centers, the Institution has facilities in eight states and the Republic of Panama. Research projects in the arts, history and science are carried out by the Smithsonian all over the world. Some of the Smithsonian's research centers include the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts, the Smithsonian Marine Station at Link Port, in Florida, and the Smithsonian Tropical Research Institute, in Panama.

For membership information of pre-visit planning material, write or call the Visitor Information and Associates Reception Center, Smithsonian Institution, Washington, D. C., 20560, (202) 357-2700 (voice), (202) 357-1729 (TTY). You may also visit the Smithsonian through our web site, www.si.edu.

HISTORY

James Smithson (1765 –1829), a British scientist, drew up his will in 1826 naming his nephew, Henry James Hungerford, as beneficiary. Smithson stipulated that, should the nephew die without heirs (as he did in 1835), the estate would go to the United States to found "at Washington, under the name of the Smithsonian Institution, an establishment for the increase and diffusion of knowledge..."

On July 1, 1836, Congress accepted the legacy bequeathed to the nation by James Smithson, and pledged the faith of the United States to the charitable trust. In 1838, following approval of the bequest by the British courts, the United States received Smithson's estate—bags of gold sovereigns—then the equivalent of $515,169. Eight years later, on August 10, 1846, an Act of Congress signed by President James K. Polk, established the Smithsonian Institution in its present form and provided for the administration of the trust, independent of the government itself, by a Board of Regents and Secretary of the Smithsonian.

SMITHSONIAN MUSEUMS, GALLERIES AND ZOOS

Smithsonian Institution
Anacostia Museum
Arthur M. Sackler Gallery
Arts and Industries Building
Cooper-Hewitt, National Design Museum
Freer Gallery of Art
Hirshhorn Museum and Sculpture Garden
National Air and Space Museum
National Museum of African Art
National Museum of American History, Behring Center
National Museum of the American Indian
National Museum of Natural History
National Portrait Gallery
National Postal Museum
National Zoological Park
Renwick Gallery
S. Dillon Ripley Center
Smithsonian American Art Museum

⊛ SMITHSONIAN®

CRYSTAL RADIO™

WARNING!
ONLY FOR USE BY CHILDREN OVER 8 YEARS OLD.

READ THE INSTRUCTIONS BEFORE USE, FOLLOW THEM AND KEEP THEM FOR REFERENCE. STORE THE SET OUT OF REACH OF SMALL CHILDREN.

PLEASE KEEP A NOTE OF OUR NAME AND ADDRESS DETAILS FOR FUTURE REFERENCE. IN EUROPE CONTACT:

CE

NSI SIMM GAmbH
D 76162 KARLSRUHE
GERMANY
49-0721-9584-116

PLEASE BE SURE TO READ THE ADVISE FOR SUPERVISING ADULTS AND THE SAFETY RULES CONTAINED IN THIS BOOKLET.

ADVICE FOR SUPERVISING ADULTS:

• READ AND FOLLOW THESE SAFETY INSTRUCTIONS,
THE SAFETY RULES AND KEEP THEM FOR REFERENCE.

• SUPERVISING ADULTS SHOULD DISCUSS ANY WARNINGS
AND SAFETY INFORMATION WITH THE CHILD BEFORE
COMMENCING THE ACTIVITIES.

SAFETY RULES:

• DO READ THE INSTRUCTIONS BEFORE USE, FOLLOW THEM
AND KEEP THEM FOR REFERENCE.
• DO KEEP YOUNG CHILDREN AND ANIMALS AWAY FROM THE
ACTIVITY AREA.
• DO STORE THE SET OUT OF REACH OF YOUNG CHILDREN.
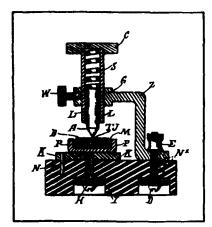
Facts about early crystal development
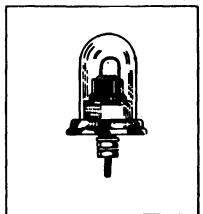
Several inventors receive credit for developing a "crystal detector", a device which can pass current better in one direction than the other in an electrical circuit. In Germany, in the 1870's, Karl F. Braun noticed this property in certain mineral substances. Commercial development required a technology that would be suitable. This came with the advent of wireless communication and radio after 1900.
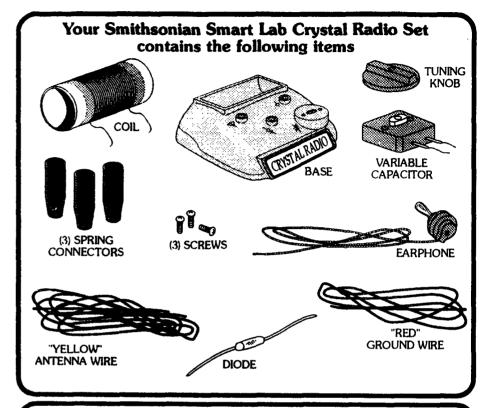
In Japan, Wichi Torikata investigated many minerals, including zincite, pyrolusite, iron pyrites and galena. In the United States, Greenleaf W. Pickard, associated with the Wireless Specialty Apparatus Co. of Boston, conducted extensive experiments as well. He became well known for the PERIKON Detector which employed zincite and chalcopyrite. The early 20th century work was generally done between 1900 and 1912. Included among the scientists is General H.H.C. Dunwoody of the U.S.Army who developed a carborundum detector.
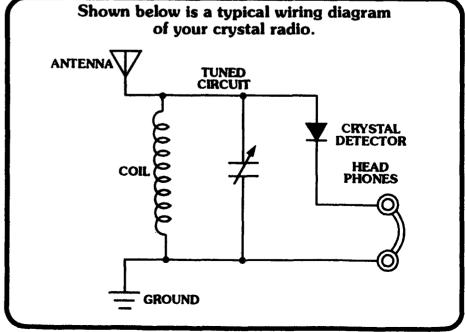
Until radio broadcasting began in 1920 (Station KDKA Pittsburgh) experimenters utilized the crystal detector for radiotelephone communications and for reception of certain Morse Code signals. There was also extensive use by the U.S.Navy and other maritime services.

The modern crystal diode, such as provided with your set, is a spin-off of radar technology developed during World War II. The basic principle, however, remains the same..i.e. to remove the AUDIO (speech, music or Morse-Code) from the radio frequency carrier wave. In one early Pickard crystal design (left) a stiff metal point is adjustable over the crystal surface. In a later version a "Catswhisker" spring impinged on a piece of galena (right) and was varied to give a louder signal.

Your Smithsonian Smart Lab Crystal Radio Set contains the following items

COIL

TUNING KNOB

VARIABLE CAPACITOR

BASE

(3) SPRING CONNECTORS

(3) SCREWS

EARPHONE

"YELLOW" ANTENNA WIRE

DIODE

"RED" GROUND WIRE

Shown below is a typical wiring diagram of your crystal radio.

ANTENNA

TUNED CIRCUIT

CRYSTAL DETECTOR

COIL

HEAD PHONES
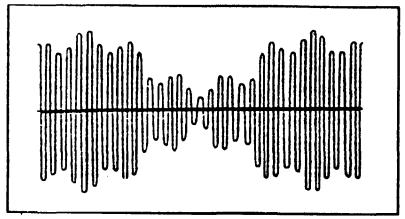
GROUND

When these two different energy waves are put together by the mixer stage of the transmitter, the resulting wave, which is the signal that the station broadcasts, looks like this:

AM Broadcast Wave

The signal is called a MODULATED wave. You will notice that each of the waves is the same length, but the heights vary. Since the height of a wave is called its AMPLITUDE. The type of transmission that WXYZ uses is called AMPLITUDE MODULATION. That's why WXYZ is called an AM station.

The crystal radio that you built works in just the opposite way as the radio station. The modulated signal broadcast by WXYZ is "picked up" by the antenna on your radio.
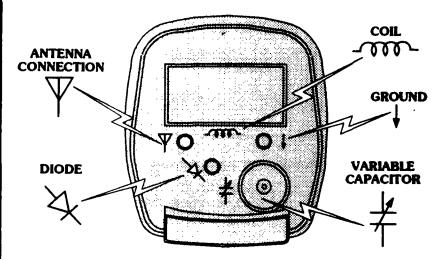
The antenna is connected to the tuner circuit of the radio, which is made up of a coil of wire and a variable capacitor connected together. As you move the tuning knob, you are able to select, or isolate, the particular

This signal is then DEMODULATED by the diode, separating the voice (audio) wave from the carrier wave.

The voice (audio) wave then moves to the earphone, where it is changed back into sound waves that you can hear.

Before starting to assemble your Smart Lab Crystal Radio please learn the basic electronic symbols that are shown on the plastic base.

Explanation of components

Variable Capacitor - It is used to tune the radio to a station. The leads that are soldered to the lugs are used to connect it to the circuit.

Diode - A small crystal is sealed inside with leads connected to it. Pay careful attention to the position of the black bands painted on the diode close to one end.

Coil- This is a radio-tuning COIL. It was made by winding enameled copper wire around a paper core 80 times. The leads have been stripped and tinned so it can be connected to the circuit.

Earphone - It contains a small crystal that can make enough electricity to drive a metal diaphragm to produce sound. The leads have been stripped and tinned so it can be connected to the circuit.

Antenna - This is a wire used for radiating or receiving radio waves.

Ground - This is a wire used to make a electric connection with the earth or other type of grounding source to create a common return for an electric circuit.
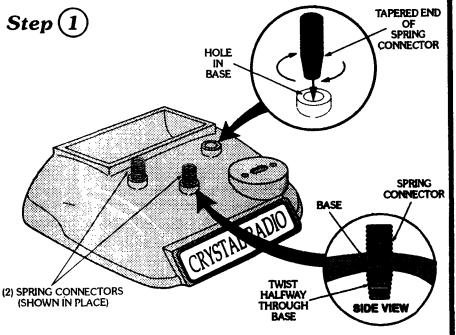
How to assemble your Crystal Radio

Before starting, it's important to assemble the parts correctly.
Carefully read the following instructions step by step
and have fun . . . half the fun is knowing you made
your very own crystal radio.

Step (1)

TAPERED END
OF
SPRING
CONNECTOR

HOLE
IN
BASE

SPRING
CONNECTOR

BASE

TWIST
HALFWAY
THROUGH
BASE

SIDE VIEW

(2) SPRING CONNECTORS
(SHOWN IN PLACE)

A. Locate the three Spring Connectors.
B. Insert one of the Spring Connectors (tapered end first) into one of the three holes on the Base.
C. Push down on the Spring Connector and twist to the right until the Connector is approximately half way through the base.

NOTE: Check to see if the Spring Connecter is inserted halfway through the Base by turning Base upside down (see insert).
D. Insert the remaining two Spring Connectors into the holes on Base the same way.

Radio Technology
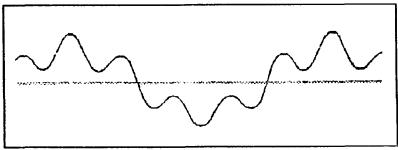
Now that you have your crystal radio assembled and working,
it is time to take a brief look at how it works. To do this
you are going to take an imaginary trip to a
radio station WXYZ, in anytown,USA.

When you arrive at WXYZ, you are met by the station's general manager, Mr. Smith, who is going to show you around.

First, Mr. Smith takes you to a STUDIO, a special room that a program comes from. There is a lot of equipment, including dials and switches, record players, lots and lots of records, and microphones.
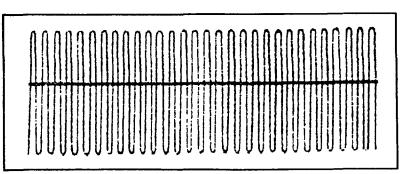
Mr. Smith tells you that when you talk into one of the microphones, your words go into the station's electronic equipment. There, it is mixed with the station's carrier wave, and sent out into the air through the station's transmitting antenna - that tall tower you saw on top of the radio station building.

Then Mr. Smith takes a piece of paper and a pencil, and draws diagrams to show you what happens. He begins by saying that all energy travels in waves, and since the sound we make when we talk is a form of energy, it might look something like this:

Sound Wave

The station's carrier wave is a radio wave, which is also a form of energy, which looks something like this:

Radio Carrier Wave

After you have attached all the wires, carefully read the following instructions to make sure your crystal radio operates properly.

A. Place the earphone in your ear and turn the Tuning Knob clockwise and counterclockwise until you pick up the strongest signal.
Note: If you do not pick up a strong signal or do not hear anything, please read some of the troubleshooting solutions explained below.

TUNING KNOB
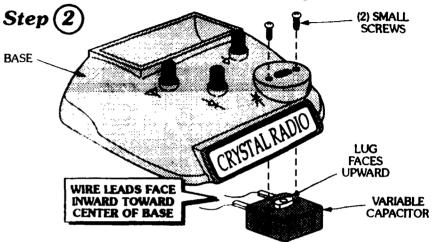
TROUBLESHOOTING HINTS

1. There may be a bad connection caused by improper assembly of your radio. Check all the spring connectors to make sure the wires are attached correctly. If a wire is loose or the bare wire is not making good contact with the spring connector, the radio may not work.

2. If the uninsulated part of the Antenna wire touches anything other than the Antenna Connection on the radio, the radio may work improperly or not at all..

3. If the bare stripped end of the Ground wire is not wrapped tightly around a water pipe (so that it makes good contact), try taping the wire to the shiny part of the pipe using duct tape. If the pipe is dull or rusty, use a piece of sandpaper to gently sand the area where the wire makes contact. Also make sure the wire doesn't touch anything other than the Ground Connection or the radio may not work properly.

4. You may live in an area where radio reception is generally poor. Instead of trying to use your radio during the day, try at night when many radio stations are received better.

5. The Antenna on your radio is a very important component. Make the Antenna as long and high above the ground as possible. If you live in a multi-story dwelling the highest floor should be used when operating the radio. Insulated bell-wire, which is available in hardware stores, makes a very good substitution for your present Antenna.
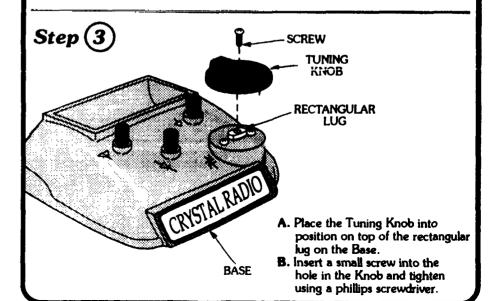
For this step you will need the variable capacitor, (3) screws and turning knob. Be carefull not to misplace the small screws when assembling.

Step ②

BASE

(2) SMALL SCREWS

LUG FACES UPWARD

VARIABLE CAPACITOR

WIRE LEADS FACE INWARD TOWARD CENTER OF BASE

A. First place the Variable Capacitor up into the base so the holes on the capacitor line up with the holes on the base.
Note: When you do this be sure the wire leads face inward toward the base as shown and lug faces upward.
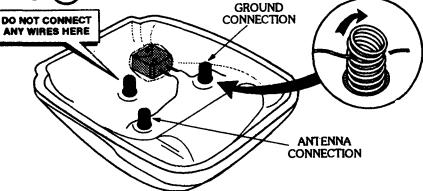
B. Next insert the two small screws into the holes on top of the Base and tighten using a small phillips screwdriver.

Step ③

SCREW

TUNING KNOB

RECTANGULAR LUG

BASE

A. Place the Tuning Knob into position on top of the rectangular lug on the Base.
B. Insert a small screw into the hole in the Knob and tighten using a phillips screwdriver.

How to assemble your Crystal Radio (cont'd)

Step ④ For this step turn the Base upside down.

DO NOT CONNECT ANY WIRES HERE

GROUND CONNECTION

ANTENNA CONNECTION
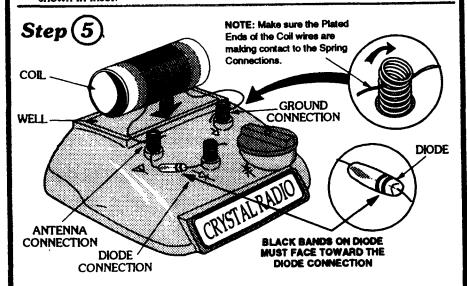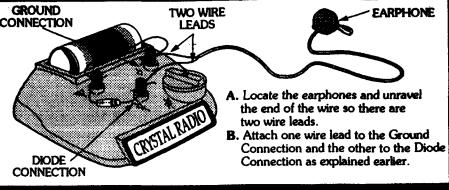
A. Turn the Base upside down.
B. Connect one wire to the Antenna spring connector by bending the spring slightly to one side and inserting the wire through as shown in inset.

C. Connect the other wire to the Ground spring connector in the same manner.
NOTE: Do not connect any of the wires to the Diode spring connector or your crystal set will not work properly.

Step ⑤

NOTE: Make sure the Plated Ends of the Coil wires are making contact to the Spring Connections.

COIL

WELL

GROUND CONNECTION

DIODE

ANTENNA CONNECTION

DIODE CONNECTION

BLACK BANDS ON DIODE MUST FACE TOWARD THE DIODE CONNECTION
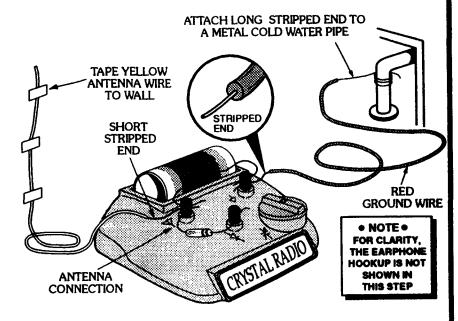
A. Next place the Coil in the well of Base with the wires hanging loosely toward the connectors .
B. Connect one of the plated ends of Coil wires to the Ground Connector by bending the Spring Connector as shown in inset. Connect the other Coil wire to the Antenna Connector.

C. Place the Diode between the Diode and Antenna Connections carefully noting the black bands face toward the Diode Connection.
D. Connect ends of Diode wires to Antenna and Diode spring connectors shown.

How to attach the earphones

GROUND CONNECTION

TWO WIRE LEADS

EARPHONE

DIODE CONNECTION

A. Locate the earphones and unravel the end of the wire so there are two wire leads.
B. Attach one wire lead to the Ground Connection and the other to the Diode Connection as explained earlier.

Final assembly and operation

ATTACH LONG STRIPPED END TO A METAL COLD WATER PIPE

TAPE YELLOW ANTENNA WIRE TO WALL

SHORT STRIPPED END

STRIPPED END

RED GROUND WIRE

ANTENNA CONNECTION

● NOTE ●
FOR CLARITY, THE EARPHONE HOOKUP IS NOT SHOWN IN THIS STEP

A. Locate the Red Ground and Yellow Antenna wire.
B. Attach the stripped end of Yellow Antenna wire to the Antenna Connection on the Base.
Note: The Antenna should be strung in a straight line away from power lines and large metal objects. We suggest taping it high up on a wall using masking tape as shown.

C. Next attach the short stripped end of the Ground wire to the Ground Connection.
D. Attach the Long stripped end to a metal cold water pipe, metal radiator pipe, etc., near where the radio will be used. Have a adult help you with this since some pipes may be hot.
Note: If there is not enough stripped, bare wire to wrap around a pipe then ask an adult to strip more insulation off the end using a pair of wire strippers.

Germanium vs Silicon Diode Testing: Read this document carefully, so you will not be the victim of cheap knock-off or the wrong type diodes.

The general rule is that **<u>SILICON</u>** diodes have a voltage drop across the Anode to Cathode of 0.7 V (7/10 tenths), and the **<u>GERMANIUM</u>** diodes have a voltage drop of 0.3 V (3/10 tenths) more or less. Either diode voltage drop (silicon or germanium) will display a reading within approximately 5% of these readings.

Most digital meters (DVM) have a switch setting ⊶ that is used to measure voltage drop across these diodes. This setting is usually indicated by a diode symbol to let the user know the DVM is capable of measuring forward bias voltage. This setting will tell you immediately if the diode is a **germanium**, or **silicon** diode. You need to set the selector switch on your meter to the diode test symbol ⊶.

Measuring Forward Bias Voltage

To measure the forward bias voltage characteristic you connect the black probe of your meter to the cathode terminal. The cathode terminal is on the end with a band. You then connect the red probe to the anode terminal.

Set your DVM/DMM to the *diode test mode* ⊶, it should provide you with the respective voltage drop. If the figure is 0.3 V or less, the diode is a *germanium* type. If the voltage drop is 0.7 V or less the diode is a *silicon* diode.

Sensitivity and Forward Bias Characteristics:
The sensitivity of a diode to radio waves depends upon its *forward bias voltage*. This is the voltage across the diode terminals. When it falls below this threshold value, the diode will stop conducting. Obviously, the lower this threshold value is, the greater the sensitivity of the diode to the weak radio signals.

CONCLUSIONS:
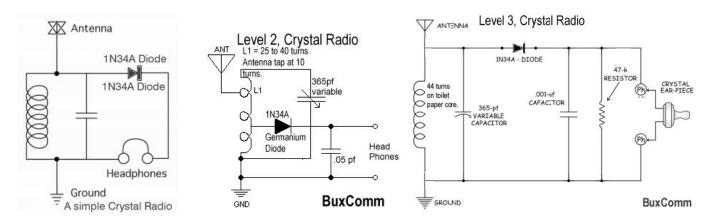
When comparing germanium diodes with silicon diodes of similar forward bias voltage, silicon diodes do not perform as well as the germanium.

Germanium has many properties that silicon diodes do not have. Germanium requires very little forward current. Forward current in a germanium diode is in the *micro*ampere region, while silicon diodes require *amperes*. This makes *germanium* a much better choice for both medium and high frequency radio signals.

Germanium also exhibits a very low, point-contact junction capacitance, while the silicon diode has much higher capacitance. A low junction capacitance allows germanium diodes to operate more effectively at high RF frequencies.
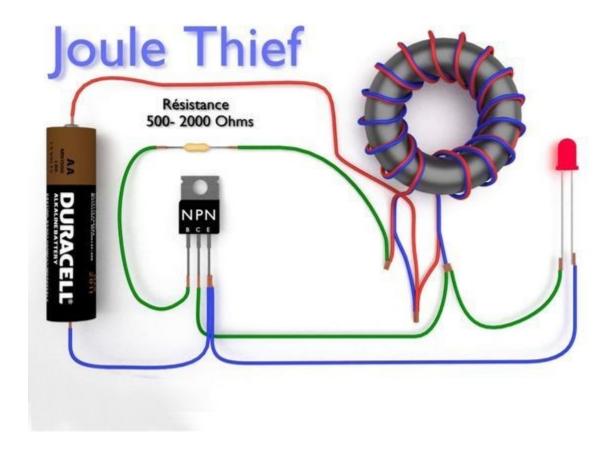
In addition, reverse leakage current for germanium diodes is in the magnitude of 1000, much more than silicon. This makes the non-linear characteristics of the germanium diode much more effective for RF detection and demodulation than silicon.
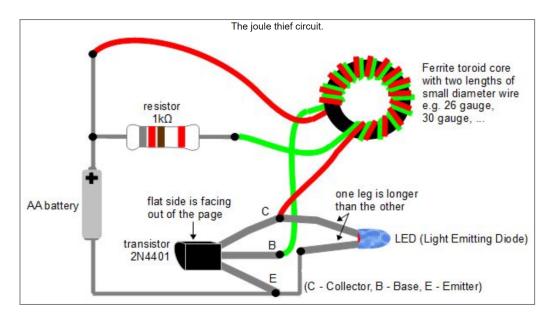
Therefore, our conclusions are; germanium diodes provide the best performance in crystal radios, RF probes, HF, and VHF signal detection.

BuxComm High Quality, CRYSTAL RADIO ANTENNA KIT, $14.95 cat# 50CRAK

This Crystal Radio Antenna kit consists of 50 feet of #16 AWG insulated & stranded copper antenna wire, two heavy-duty, Delrin, UV resistant antenna insulators and instruction sheet.

Joule Thief

Résistance
500- 2000 Ohms

DURACELL ALKALINE BATTERY
AA

NPN
B C E

The joule thief circuit.

Ferrite toroid core with two lengths of small diameter wire e.g. 26 gauge, 30 gauge, ...

resistor
1kΩ

AA battery

flat side is facing out of the page

transistor
2N4401

one leg is longer than the other

C

B

E

LED (Light Emitting Diode)

(C - Collector, B - Base, E - Emitter)

Transistor - The legs of the transistor can be determined by noticing that there's a flat side to the transistor case. See the diagram above. A large number of transistors have been reported to work: 2N4401, NET123AP, BC547B, 2SC2500, BC337, PN2222, to name just a few.
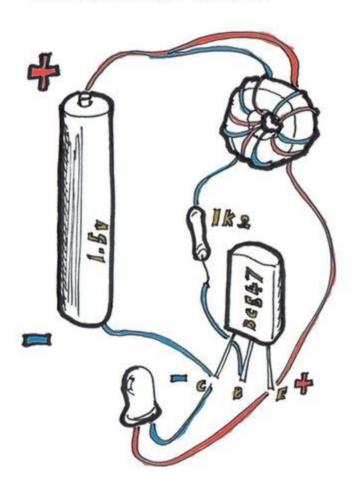
LED - One leg of the LED is longer than the other leg. Use this to determine which one goes where. See the diagram above.
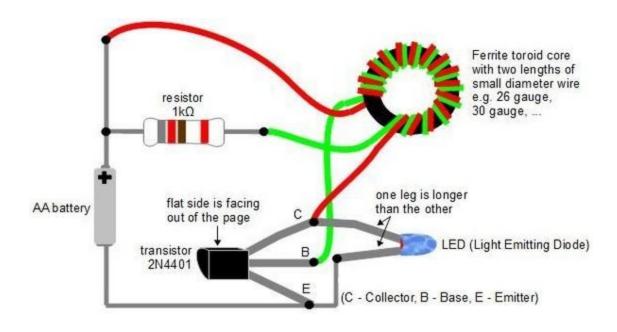
Resistor - The diagram says use a 1 kilo ohm resistor but I've used an 820 ohm one just fine. I've also seen a 2 kilo ohm one in use. Use whatever works for you. You can also use a potentiometer (a variable resistor) so that you can easily adjust it to select the resistance that gives the best light.

Toroid ferrite core - Some people have gotten these by opening up compact fluorescent lightbulbs (CFLs). I took mine out of some device whose original function I don't know. To get it working, my first one had just 13 turns for each wire and I used a 30 gauge wire and a 26 gauge wire. The wire must be insulated. A variety of number of turns will work. This is something you can play with. Look at the diagram carefully to determine where the wires connect to.

**Its a Rubbish Challenge
Dog Light**
Joule Thief
electronics circuit

resistor
1kΩ

Ferrite toroid core
with two lengths of
small diameter wire
e.g. 26 gauge,
30 gauge, ...

AA battery

flat side is facing
out of the page

one leg is longer
than the other

C

transistor
2N4401

B

LED (Light Emitting Diode)

E

(C - Collector, B - Base, E - Emitter)

MAKING A SIMPLE JOULE THIEF (MADE EASY)
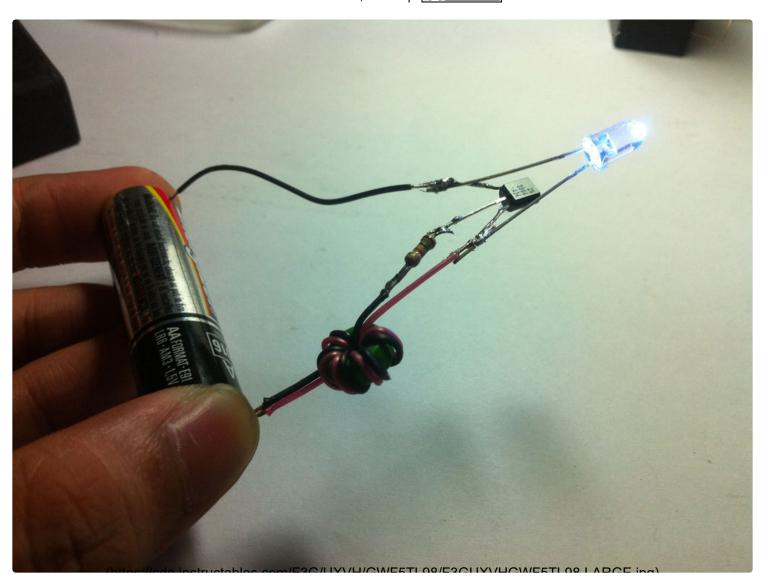
Technology (/technology/) > Leds | by ASCAS (/member/ASCAS/)   Follow

195,691   ♥ 573   165

Posted Dec. 24, 2011 | (cc) BY-NC-SA

(https://cdn.instructables.com/FB5/ROY8/GWHN3 (https://cdn.instructables.com/FF9 (https://cdn.instructables.com/F3C

Today I am showing you how to make a very simple joule thief. A joule thief has many applications, the best gadget that I made with was a "Water Powered Lamp", soon I'm going to post on a guide about it but first I need to post this guide. I used an iPhone 4S as my camera :)))
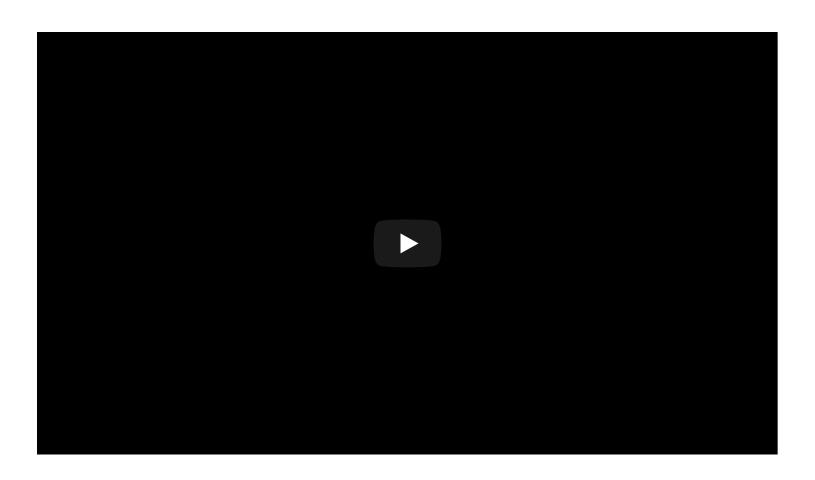
What Is A Joule Thief ?

To simplify everything, a "joule thief" is a circuit that helps drive an LED light even though your power supply is low. What can we do with it? We can use it to squeeze the life out of our old, almost drained, non functioning batteries. This project can also be considered as a green and environmental experiment, we can also use it as a flashlight that can be ran by an old, weak, almost drained battery. I even tried to use my water powered battery from my previous instructable the "Water Powered Calculator (https://www.instructables.com/id/Water-Powered-Calculator/)", the project was featured and displayed in instructable's front page in the "Technologies" category.
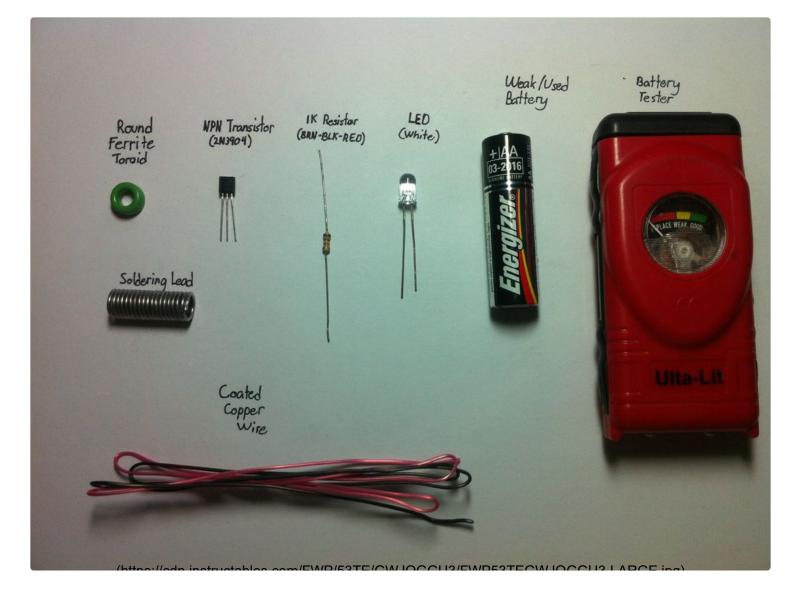
My Next Projects That Involves A Joule Thief: (soon to be posted)
- Water Powered Lamp
- Water Powered Flash Light
- Dead Battery Drainer Lamp

Here's A Video From Make Magazine:

Add Tip   Ask Question

Step 1: Parts and Materials
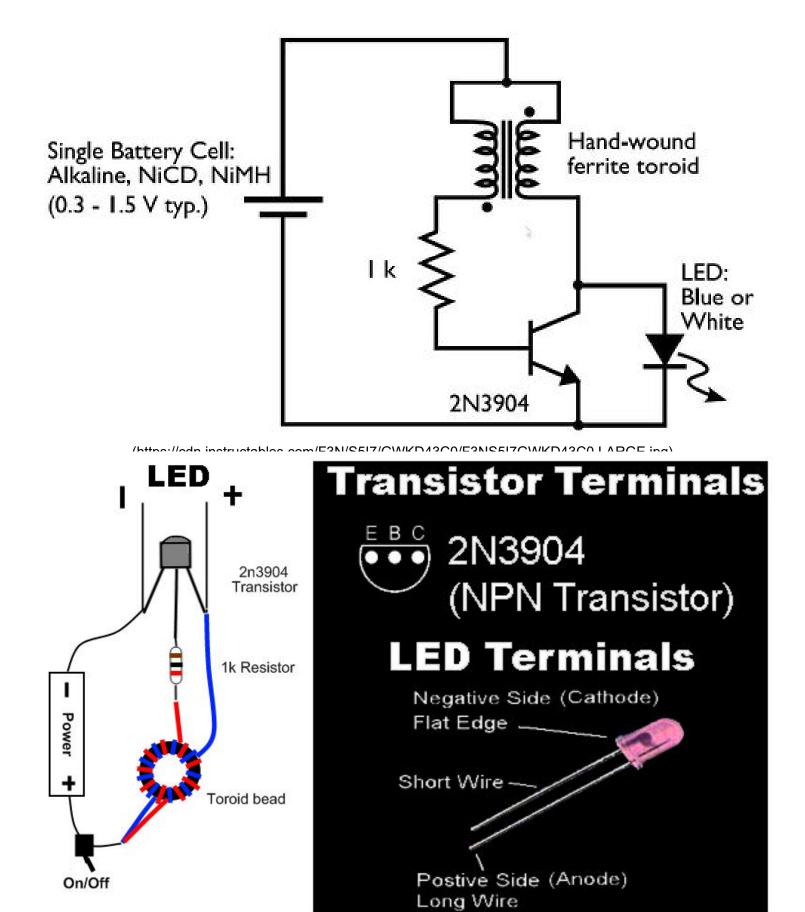
The Parts Needed Are: (click the item to know where to find/ buy)

- Round Ferrite Toroid (can be found in old CFL bulbs)
- Old/ Used Batteries (can be found in garbage cans)
- NPN Transistor (2N3904) (http://www.radioshack.com/product/index.jsp?productId=2062609)
- 1K Resistor (BRN-BLK-RED) (http://www.radioshack.com/product/index.jsp?productId=2062343)
- LED Light (http://www.radioshack.com/product/index.jsp?productId=3096133)
- Battery Tester (http://www.tooldistrict.com/Ulta-Lit-Battery-Tester-5001-p/901233.htm) (optional)
- Soldering Lead (http://www.radioshack.com/product/index.jsp?productId=2062717)

- [Copper Wire/ Magnet Wire (http://www.radioshack.com/product/index.jsp?productId=2036277)](http://www.radioshack.com/product/index.jsp?productId=2036277)
- [Battery Case/ Holder (http://www.radioshack.com/product/index.jsp?productId=2062247)](http://www.radioshack.com/product/index.jsp?productId=2062247)

I want to share something. Here in the Philippines electronic parts are extremely cheap, they are extremely far cheaper from
radio shack, for example one transistor costs (2 phil. pesos - 6 US cents), a LED cost (9 phil. peso - 29 US cents) and a 1K resistor cost (25 phil. cents - 0.8 US cents). I usually buy thing from Deeco or Alexan. Usually prices here are 15x cheaper from radio shack. Price conversion - $1 US Dollar = P0.31 Philippine Peso (12/24/11).

Add Tip   Ask Question

Step 2: Schematic Diagrams

Single Battery Cell:
Alkaline, NiCD, NiMH
(0.3 - 1.5 V typ.)

1 k

Hand-wound
ferrite toroid

LED:
Blue or
White

2N3904

LED

−   +

2n3904
Transistor

1k Resistor

Power
−
+

Toroid bead

On/Off

Transistor Terminals

E B C

2N3904
(NPN Transistor)

LED Terminals

Negative Side (Cathode)
Flat Edge

Short Wire

Postive Side (Anode)
Long Wire

Here are the schematic diagrams that are involved with the joule thief circuit.

Step 3: Winding Wire at the Toroid

(https://cdn.instructables.com/FLD/JSNX/GWF5TL99/FLDJSNXGWF5TL99.LARGE.jpg)
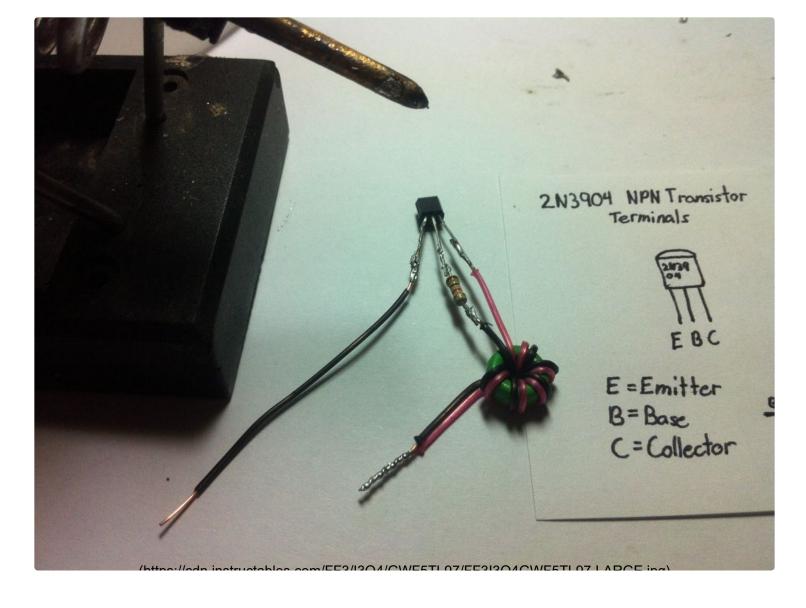
First, connect both ends of the copper wire before wounding, be sure to remove the insulation. Then try to solder the ends so it would not split up. Second, wind the wire until you run out of space in the round ferrite toroid. I have some tips for you, try to use a gauge #22 enamel coated copper wire for better performance, oh! my last tip is that "the more you wind the wire to the ferrite toroid the better".

Add Tip   Ask Question

Step 4: Soldering the Resistor
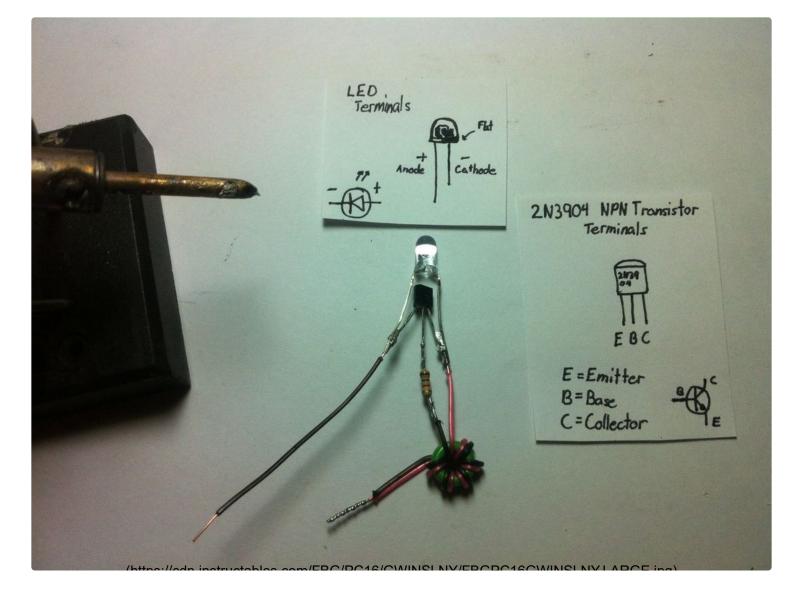
Solder the resistor with one end of the wounded ferrite toroid's wire. Oh! also don't forget to level the other end of the resistor with the other unused wire from the wounded ferrite toroid.

Add Tip   Ask Question

Step 5: Soldering the Transistor

Solder the proper connections to the transistor. For the emitter - connect another wire, the wire will be connected to the negative part of the battery. For the base - solder the other end of the resistor to the base. For the Collector Solder the unused wire of the ferrite toroid.

Add Tip   Ask Question

Step 6: Soldering the LED

(https://cdn.instructables.com/FRC/RC16/CWINSLNX/FRCRC16CWINSLNXLARGE.jpg)

Solder the shorter wire of the LED to the tansistor's emitter and the longer part of the LED to the transistor's collector. After all that, you can now trim the excess wires.

Add Tip   Ask Question

Step 7: Time to Look for Old Batteries
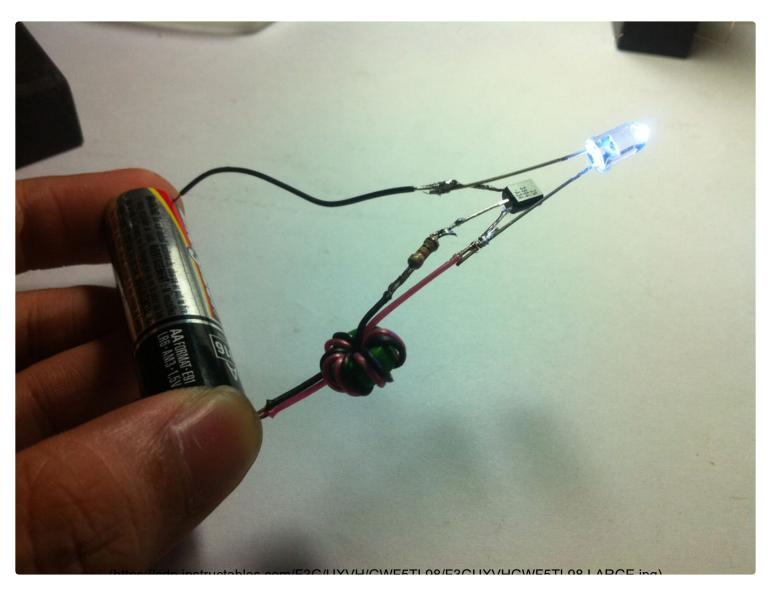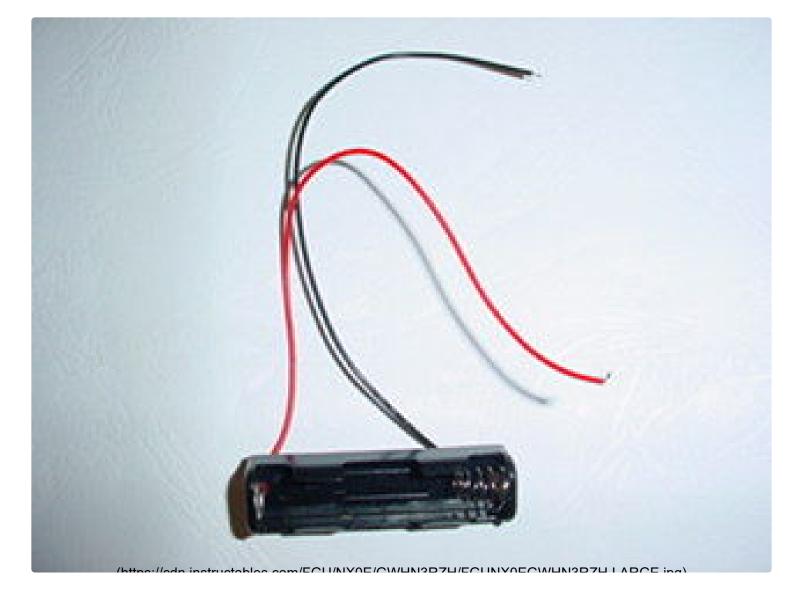
Use your battery tester to confirm that your battery is close to death. The tester is only an optional tool, it's just used to determine the battery's remaining power.

Step 8: Time to Test It - You're Done !!

(https://cdn.instructables.com/FCU/NX0F/GWHN3BZH/FCUNX0FGWHN3BZH.LARGE.jpg)

The wire connected to the transistor's emitter should be connected to the battery's negative side and the remaining wire of the ferrite toroid should be connected to the battery's positive side. Oh! one more thing, I advise everyone to use a battery case or attach a conductive magnet for each wire, so you wouldn't hold it all the time.

Add Tip   Ask Question
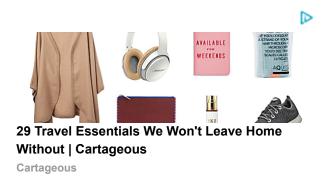
8 People Made This Project!

"Joule Thief" LED Night Light

by **ledartist** on September 2, 2011

Table of Contents

Author:ledartist   My Blog
I work with LEDs and microcontrollers to create beautiful objects.

Intro: "Joule Thief" LED Night Light

I have many used batteries around. Remote controls, cameras, many electronic gadgets all use batteries, mostly AA size. I always felt guilty for throwing away the used batteries. I know there are rechargeable batteries, but many electronics don't work well with rechargeables.
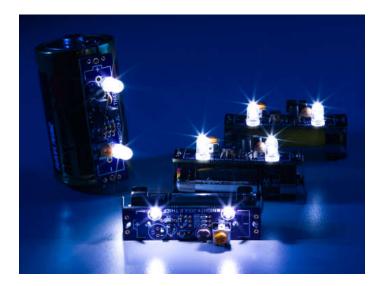
I also know that those "used" or "spent" batteries usually have some juice left in them. So to come up with a good use of used batteries, I've created a LED night light.

I like having a little night light on when I sleep. LEDs are perfect for this purpose, because they are energy efficient, and good at providing low intensity illumination.

This LED night light operates with just one battery. It utilizes a little circuit called Joule Thief to boost voltage out of an AA battery. I also added a light sensor to turn it on automatically when the surrounding is dark.

The circuit is energy efficient, and requires very low voltage to work. So it effectively sucks every bit of energy out of batteries. This type of circuit is often called "Joule Thief", because it works as though stealing every bit of energy (Joule is a unit for energy) out of battery.
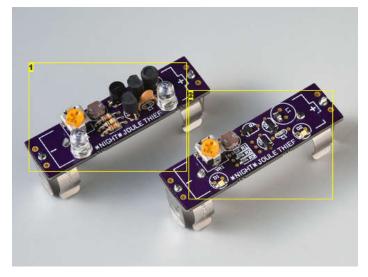
I'm calling this project Night Joule Thief.

Step 1: Features
Here are the highlights of the Night Joule Thief.

- Compact & streamlined design
- Uses only one AA battery (or any 1.5V battery you can hook up to)
- Easily adaptable to different size batteries - hook up holes to attach home made clips
- Two white LEDs
- Automatic turn on via a light sensor (adjustable sensitivity level)
- Energy efficient - works even with a run-down battery, down to 0.6V
- Choice of through-hole only components or SMD - mix & match on the same PCB
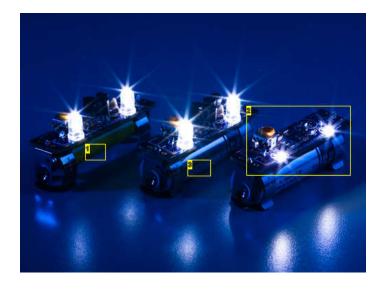
http://www.instructables.com/id/Joule-Thief-LED-Night-Light/

Image Notes
1. Through-hole version.
2. SMD (Surface Mount Device) version. Hard to believe this is the same circuit,
but it works exactly the same as one on left. Ok, the LEDs are slightly less bright.

Image Notes
1. AAA battery inserted with a help of small magnets.
2. SMD version - very low profile.
3. Standard AA battery snugly fit in the clips.

Image Notes
1. Other sizes, such as D cell can be attached by making a pair of little metal clips.
Just cut and bend paper clips.

Step 2: Technical Overview

"Joule Thief" circuit is an inductor based voltage booster circuit to light LEDs with low supply voltage. As most of you know LEDs need higher than 2V (3V for white LEDs), so usually at least two batteries are needed to light them. The "Joule Thief" circuit was published in 1999 and has been quite popular. You can see the principle of the circuit here. http://en.wikipedia.org/wiki/Joule_thief

My version is a variation that uses single coil inductor, to make the inductor easily obtainable. I design the circuit using readily available parts only, to make it an ideal DIY project.

Circuit
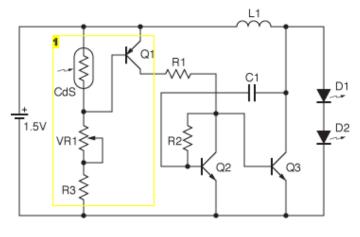
The L1, Q2, Q3, C1, R2, and LEDs D1 & D2 make the Joule Thief. And the Q1, and the rest of the parts form the light sensor. CdS is the device that actually senses the light and change its resistance accordingly. When the surrounding of CdS is bright, it has low resistance (anywhere around 1k to 3k ohm), and when the surrounding is dark, the resistance goes up to 100k to 3M ohm range. So in this circuit, the base voltage of Q1 is controlled by the ambient light level. When the base voltage of Q1 goes more than 0.6V below the power supply(battery) voltage, current goes through R1, turning the Joule Thief circuit on.

The Joule Thief circuit is boosting the battery voltage up to over 6V to light two LEDs in series. LEDs light up with the battery voltage as low as 0.6V! Amazing!

PCB layout can be downloaded as an editable PDF, so you can etch your own board if you like. Custom 2 layer PCB and kit are available for sale as well. The 2 layer PCBs have extra front pads for SMD where possible, so you can build the same circuit with SMD parts as you wish.

Parts List
1x CdS Photoresistor (rated 3k - 0.3M ohm) (CDS1)
1x 1k ohm (R1)
1x 100k ohm (R2)
1x 10k ohm (R3)
1x 50k ohm trim pot (VR1)
1x 22pF (C1)
1x 470uH (L1) (anywhere between 22 - 470uH would work - might have to reduce the C1 value however)
1x 2N5401 or equivalent (Q1) (or just about any general purpose PNP transistor, such as PN2907, 2N3906, etc...)
2x MPSA06 or equivalent (Q2, Q3) (or just about any general purpose NPN transistor, such as PN2222A, 2N3904, 2N4400, etc...)
2x LED (D1, D2) (Just about any LEDs can be used)
2x Battery Clips
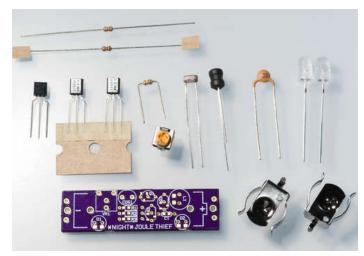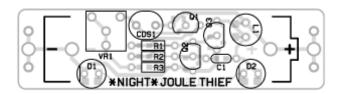
Image Notes
1. Light sensor circuit with sensitivity adjust.

File Downloads

 NightJouleThief-PCB.pdf (54 KB)
[NOTE: When saving, if you see .tmp as the file ext, rename it to 'NightJouleThief-PCB.pdf']

Step 3: **Assembly**
The assembly is very straight forward. Insert the parts into the PCB, and solder them. Start with small components, follow the order below.

Parts List (in assembly order)
1x 1k ohm (R1)
1x 100k ohm (R2)
1x 10k ohm (R3)
1x Photoresistor (rated 3k - 0.3M ohm) (CDS1)
1x 50k ohm trim pot (VR1)
1x 22pF (C1)
1x 470uH (L1)
1x 2N5401 or equivalent (Q1)
2x MPSA06 or equivalent (Q2, Q3)
2x LED (D1, D2)
2x Battery Clips
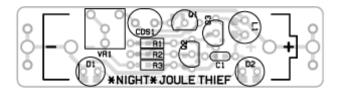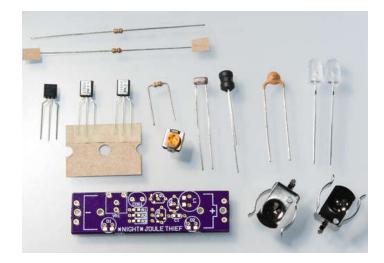
Transistors, and LEDs have polarities, so make sure to insert them in the correct orientation. Battery holders need a bit of force to snap into the holes. They attach from the back side of PCB as you can see in the picture.

Once everything is soldered in place, double check the part placement, orientation and solder joints. Then insert a battery. The polarity is marked on the front side of PCB.
If you don't see the LEDs light up, don't worry. The room is probably too bright. Take a piece of black paper or tape and block the light from hitting CdS light sensor. (and/or darken the room) If the LEDs still don't come on, turn the trimmer (the little orange thing) with a screw driver, counter clockwise. This makes the sensor less sensitive to light, so the LEDs will come on by just placing the sensor under shade, or turning off the room light.
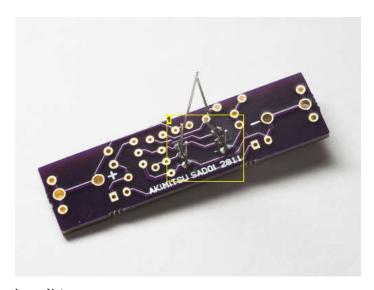
Image Notes
1. Resistors soldered in.
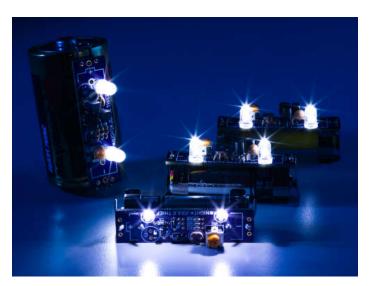
Step 4: Performance

This little night light performs very well. For starters, the brightness is not bad for using just one AA battery. I've been using these as flashlights as well.

The light sensor also works very well. Once adjusted, the light is steadily off during the day, even when you put the sensor under shade. Only when you block the sensor by a black object, the light would turn on. Yet after dusk, the light would come on when you turn off the room light.

A fresh battery lasts for weeks if only used as a night light. And the best use of this light is to "revive" used batteries. Those batteries from remote controls, cameras, etc. usually have quite a bit of juice left in them. Joule Thief sucks the juice out of those batteries till the last drop. It's like getting free energy when you can use something that were going to be thrown out.
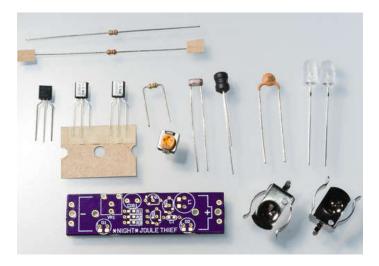
Step 5: PCB & Kit

If you are handy, you can etch your own PCB, and build this night light entirely DIY.

However, to spread the goodness of Joule Thief and to contribute to the greener earth, I am putting together the PCB & kit available.

The details can be found here: http://www.instructables.com/community/Joule-Thief-LED-Night-Light-Kit-PCB/
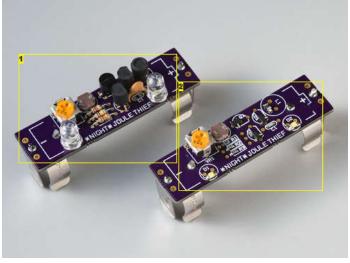
Image Notes
1. Through-hole version.
2. SMD (Surface Mount Device) version. Hard to believe this is the same circuit, but it works exactly the same as one on left. Ok, the LEDs are slightly less bright.

Related Instructables

Joule Thief - use LEDs with only one AA battery! by RazorConcepts

Ferrite, the joule thief (Photos) by botronics

Electronic Night Light by TinkerJim

Ultraviolet Light Pen by junits15

Mini 'Joule Thief' (Photos) by Mudbud

Altoids Joule Thief Flashlight (Photos) by cynical_chemical

Comments

33 comments   **Add Comment**

---

tosacj says: Sep 4, 2011. 10:47 AM **REPLY**
Great idea. Let me know when the kit becomes available.

---

ledartist says: Sep 4, 2011. 12:38 PM **REPLY**
I will post the information on "Kits" section of the forum and on my blog (theLEDart.com). The kit & PCB are scheduled to be ready in mid September.
Thanks, Aki

---

haxcess says: Sep 5, 2011. 8:16 AM **REPLY**
This is a very neat application of the joule thief. I am wondering if it would be possible to replace the light-sensor circuit with a small battery charging circuit (solar). I suppose that would ruin the beautiful form factor you have here.

---

apalacios2 says: Sep 5, 2011. 7:07 AM **REPLY**
Excellent circuit. May I remove the CdS circuit and use a button switch instead (thru the resistor connected to the transistor)?
I find great that it uses a coil instead of a toroid transformer. Kudos!

---

ledartist says: Sep 5, 2011. 7:28 AM **REPLY**

Yes, you can. In fact I made sure that you can fit a 6mm tactile switch in place of trimmer pot. You can omit the CdS, and solder a switch in place of trimmer pot.

You can also remove Q1 and put a switch there as well, but my PCB won't accommodate that.

Yeah I found having to wind my own inductor a hassle, so I designed this circuit to use off-the-shelf inductors.

Aki

---

bhvm says: Sep 4, 2011. 9:17 PM **REPLY**

Excellent build!
How many mA does this circuit give? Can we use a single 150mA power LED in place of 2x 5mm LEDs?

ledartist says: Sep 5, 2011. 7:20 AM **REPLY**

The current through the LEDs is about 20mA or less peak (it's pulsed current in about 50kHz)
So you won't get any more light by using a high power LED. By using only one LED, you do get a bit more brightness per LED though. However just a regular LED will give you the same brightness as a high power LED because you are not driving the LED with high current anyway...

It's not that you can't drive high current LED, but this circuit is designed to run with as little power as possible.

Aki

---

abbtech says: Sep 4, 2011. 7:10 PM **REPLY**

Great looking project.

ledartist says: Sep 4, 2011. 7:18 PM **REPLY**

Thanks!

---

vruiz3 says: Sep 4, 2011. 1:56 PM **REPLY**

afraid of the night??? :D

---

timotet says: Sep 4, 2011. 1:21 PM **REPLY**

great job!

---

gnafpliotis says: Sep 3, 2011. 3:06 PM **REPLY**

Can't we hack laptop's batteries that way to perform more? Is it only letting through a small amount of amps?

ledartist says: Sep 4, 2011. 1:17 PM **REPLY**

Laptops and many other "high-tech" gadget has many of inductor based boost/buck (to reduce voltage) voltage converter circuit in them. Especially LED back light screens use voltage converter to efficiently drive LEDs.
So in a way the battery life is already enhanced. (Some gadgets are better than others, of course...)

Aki

jolshefsky says: Sep 4, 2011. 6:19 AM **REPLY**

Yes, but only if you wanted to use your batteries once. Once a lithium rechargeable (like in laptops) is discharged past a certain point, it can no longer be recharged. So there is always some energy left in a laptop battery even when it says it's dead — albeit energy you can't use without ruining the battery.

stoobers says: Sep 4, 2011. 7:29 PM **REPLY**

I've heard this theory. Have you done any experiments that might validate this theory?

jamwaffles says: Sep 4, 2011. 6:14 AM **REPLY**

This might not be 100% accurate but yes, that's why we can't use the same technique for laptops. Joule thieves ramp the voltage up, but the current goes down due to V = IR, assuming the load is constant which it is if it's the same laptop ;-)

jamwaffles says: Sep 4, 2011. 6:15 AM **REPLY**

A very professional looking product. This is a great idea; the amount of power saved using old batteries instead of a herd of plug-in night lights, as well as the amount of batteries re-used is incredible.

Just out of interest, how long do these night lights last on one "averagely discharged" battery?

---

ledartist says:

Sep 4, 2011. 1:10 PM **REPLY**

The battery lasts very long, most likely much longer than you might think.
It's hard to define "averagely discharged", but I had one cell that was already at 0.7V (to low to be used with anything), and the although the LEDs were dim, they lighted for over 48 hours.
With my "dead" batteries coming out of a wireless mouse still have over 1V, I have plenty of light.

Aki

YakAttack says:

Sep 4, 2011. 8:24 AM **REPLY**

I concur!
@ledartist: Any tests on battery life so far?

---

nymgeek says:

Sep 4, 2011. 7:42 AM **REPLY**

Are those PCBs made by Laen at dorkbot pdx?

---

ledartist says:

Sep 4, 2011. 1:04 PM **REPLY**

Yup. They are very good!

---

acmefixer says:

Sep 4, 2011. 10:06 AM **REPLY**

I should have said in my previous comment that this does not mean there is anything wrong with your circuit. I think I would change Q2 and Q3 to an easier to obtain transistor such as a 2N4401 or PN2222A. Q1 could be a 2N3906 or just about any PNP transistor. If these changes are made, the resistors, especially R2, might need to be reduced. Thanks.

---

ledartist says:

Sep 4, 2011. 1:03 PM **REPLY**

Yes, you can use just about any general purpose transistors. R2 should be fine with most transistors, but 47k ohm might work better with some of them.

Aki

---

acmefixer says:

Sep 4, 2011. 10:00 AM **REPLY**

Unfortunately the link you gave to Joule Thief in Wikipedia is for a poorly written and totally inadequate definition. The authors (apparently several over time) do not have a firm understanding of how a JT works, and furthermore, they have made a mess of it. I have added comments in the discussion and some errors have been deleted, but it is still unworthy of use for a reference.

Also, by definition, the original blocking oscillator circuit later given the Joule Thief name used only a single transistor. Your circuit is not a one transistor blocking oscillator and bears little resemblance to the original JT, so I don't believe it should use the same name.

---

ledartist says:

Sep 4, 2011. 12:59 PM **REPLY**

I know the original circuit only uses one transistor, but the two conductor inductor is harder/expensive to purchase, and winding own inductors is a bit of work. I think using one extra transistor is a good trade off for not having to wind an inductor by hand. It also makes economical sense. (transistors are very cheap, so are single coil inductors.)

I did mention that my version is a _variation_ of original, which I find very often on the net. I also think that showing different ways to achieve the same result can be inspirational.

I also contacted the person who named the circuit "Joule Thief" (Big Clive) and he did not have a problem with me using the name.

Aki

StoryAddict says:

Sep 4, 2011. 12:54 PM **REPLY**

"Joule Thief" - I love a clever pun!

MikeDel says:

Sep 4, 2011. 10:21 AM **REPLY**

Looks good. I'm interested!

Admiral Aaron Ravensdale says:

Sep 4, 2011. 7:30 AM **REPLY**

I like your "green" thinking.
I also use a joule thief in my expedition light.

I am very glad that you will provide the parts as a kit because many kits are able to solder a circuit but etching a little bit too dangerous for them.

Thanks for your endeavour to make this instructables buyable for many people…

---

kevinhannan says:
fantasic project and incredibly well explained.

Sep 4, 2011. 6:35 AM **REPLY**

UOS says:
Great instructable! Waiting the kit version to buy one but also I'm gonna make mine from scratch :D

Sep 3, 2011. 2:59 PM **REPLY**

tphillips2 says:
You are so inventive! Cool idea and great looking light!

Sep 3, 2011. 8:07 AM **REPLY**

ledartist says:
Thanks!

Sep 3, 2011. 8:18 AM **REPLY**

gweeds says:
What a great idea! must build one and see how long it gets out of a "used'ish" battery

Sep 3, 2011. 4:41 AM **REPLY**

instructables (/)

Let's Make ...

Featured (/featured/)

Write an Instructable (/id/account/login/) | Sign Up (/account/gopro)

Classes (/classes/)   Contests (/contest/)   Forums (/community/?categoryGroup=all&category=all)   Answers (/tag/type-question/?sort=RECENT)   AUTODESK (http://www.autodesk.com)
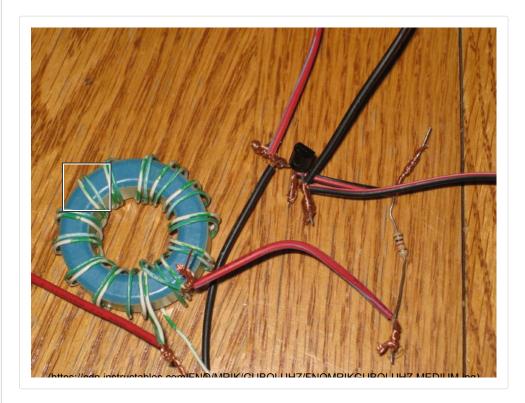
Teachers (/teachers/)   Holiday Gifts (/gifts/)

Secrets from the Sound Suite with St. Vincent
Sponsored by: **W Hotels**

(https://cdn.instructables.com/ENO/MRIK/GUROLUHZ/ENOMRIKGUROLUHZ.MEDIUM.jpg)

this is my version of the famous joule thief. there's alot of these things out there so i did alot of research and made it as simple as i can without soldering or complicated math. i harvested these parts from an older dell computer that was given to me to scrap. there is only a few parts needed to build this project:

1: toroid bead (ferrite core)
2: 1k ohm resister (brown black red)
3: npn transistor (i used the 2n3904)
4: thin wire

About This Instructable

👁 **7,340** views

❤ **23** favorites

Step 1: Wind the Coil

Joule Thief Without Soldering by

florlayamp (/member/florlayamp/) in electronics (/technology/electronics/)

Download   (/id/Joule-thief/)   4 Steps   + Collection   I Made It   Favorite   Share
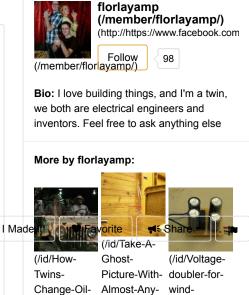
i wound a couple sizes of these things using the same wire. the type of wire i used came from an ethernet cat5 cable. i like it because its a solid core wire that stays in place where i put it. i have 11 turns of wire in the orange one and 13 in the green one. i looked all over the internet to find the reason for the number of turns but couldnt find an exact answer, and 11 winds is all the smaller (orange one) could take, and the green one i just put the 13 turns in it from the same length of wire. the orange coil is half the size (about the size of a dime) of the green one (little bit smaller than a 50cent coin).
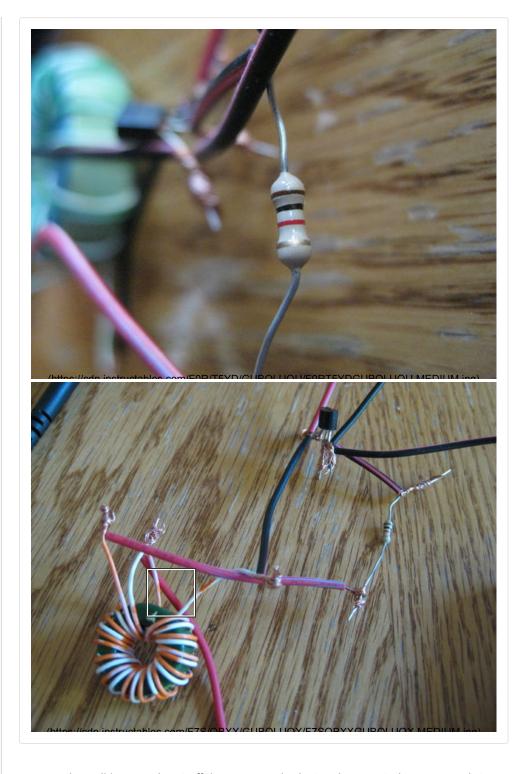
Step 2: Connect the Resistor

(https://cdn.instructables.com/F0B/T5YD/GUROLUOU/F0BT5YDGUROLUOU.MEDIUM.jpg)

(https://cdn.instructables.com/F7S/QBYY/GUROLUOY/F7SQBYYGUROLUOY.MEDIUM.jpg)
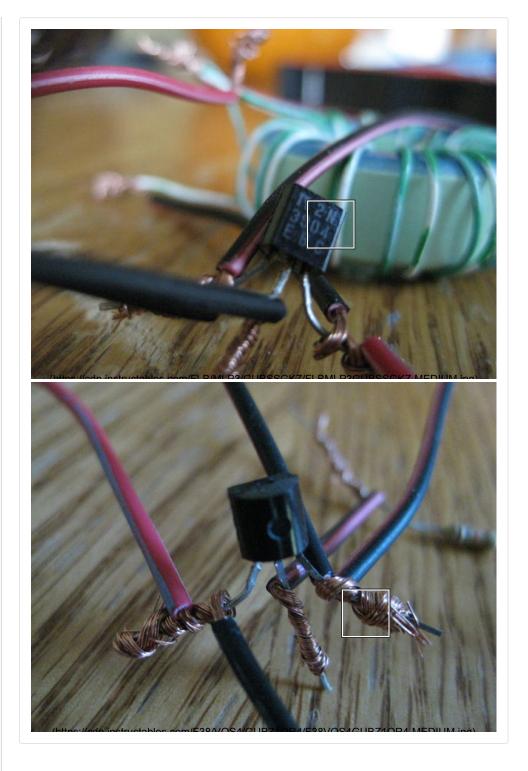
once the coil is wound, cut off the excess wire but make sure to leave enough to work with. strip the ends of all four wires, then take 1 wire of **opposite color from each side** and twist them together. then since i didnt solder any wire, i cut up several small wires to piece the rest of it together. unless you want to solder it you can skip this part. but i took a piece of wire and attached it to one of the single wires on the coil (it doesnt matter which one) and attach the other end of the piece of wire to the 1k ohm resister, then take another piece of wire and attach it to the other single wire which will go to the **collector** leg of the transistor. the final piece of wire for this step attaches to the twisted pair of wires which will become the positive input for the circuit.

Step 3: Connect the Transistor

(https://cdn.instructables.com/FLB/MLR3/GUR8SCKZ/FLBMLR3GUR8SCKZ.MEDIUM.jpg)

(https://cdn.instructables.com/F38/VOS4/GUR71QR4/F38VOS4GUR71QR4.MEDIUM.jpg)

next take a piece of wire and attach it to the other side of the resistor, and then take the other end of the piece of wire and attach it to the **base leg** of the transistor. next, as i mentioned in the previous step, take the open piece of wire from the coil and attach it to the **collector leg** of the transistor, also i attached another piece of wire to this leg for the positive output side . last of all connect two wires to the **emitter leg** which will be the negative side of the circuit. (one wire to negative on battery, then one to the negative output side of an LED light or whatever else you put on it).

Step 4: Hook It Up and Enjoy!

time to hook it up! the positive wire from the double twisted wire on the coil goes to the positive power source, and one of the negative wires on the transistor go to the negative side on the power source. the positive wire on the transistor and the other negative wire on the transistor go to the output power. once its all wired up its ready for use. as you can see in the picture, i hooked up a new AA battery and got about 5 volts! not bad for a 1.5v to 5v boost. and the orange coil circuit i also made produces about 2.5v from a 1.5v source, so this orange one only produces another volt while the green one produces alot more. i dont know why this is really. my only guess is the size of the toroid coil and the 2 winding difference. other than that its all the same parts. below is the results of some playing around with it.

green coil circuit:
1: 1.5v AA produces 5 volts
2: 2 AA 3v produces 20 volts
3: 3v input can power a 12v .25a computer fan.
4: 1.5v lights 3 green LED lights.

orange coil circuit
1: 1.5v AA produces 2.5 volts
2: 2 AA 3v produces 8 volts
3: 3v input also powers same 12v computer fan for 5 seconds...
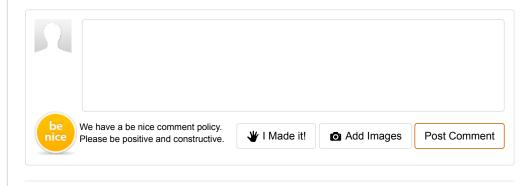
very inconsistent results between the two but thats what is going on. thanks for looking and hope this helps with your project.

Comments

✋ I Made it!   📷 Add Images   Post Comment

ToXiCATOM (/member/ToXiCATOM/)   2014-02-01   Reply

can i use 2n2222 transistor?

kyooby (/member/kyooby/) ▸ ToXiCATOM (/member/ToXiCATOM/)   Reply
2014-03-27

2N3904 = GP NPN Bi-Polar transistor commonly found in toys and small radios. Otherwise you can get them at radio shack, fry's electronics or order them online.

Other compatible ones are:

NPN

2N2222

2N2222A

2N1613

BC107

2N1711

2N2369

2N2484

2N3704

2N3705

2N3706

2N3903

2N3904

Also, if you use a 2N4401 or BC337 transistor, your LED will be brighter because they can handle more amps.

You can get the toroid and transistor from a dead CFL; the transistor is usually labeled 13002
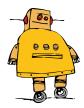
florlayamp (/member/florlayamp/) ▶ ToXiCATOM (/member/ToXiCATOM/)   Reply

2014-02-01

I tried that transistor also and I couldnt get it to output anything,
I think the internal circuits inside it are backwards for a joule thief.

⬇ More Comments

WONDER HOW TO

MAD SCIENCE

WONDERHOWTO   GADGET HACKS   NEXT REALITY   NULL BYTE

EXPERT **SERVICE.** UNBEATABLE **PRICE.**

FOLLOW US

H O W   T O

Make a "Joule Thief" and Create Zombie Batteries for More Power After Death

BY **WILLIAM FINUCANE**   ⏱ 08/23/2013 5:22 PM   `MAD POWER!`

Just about every household gadget we own runs on 1.5 volt batteries of one size or another. Wouldn't it be great if you could reuse all of those dead AA, AAA, and D batteries after they've passed on? It turns out you can make a simple circuit called a "Joule Thief" to reanimate the undead flesh of your deceased batteries and create a zombie battery.

Materials
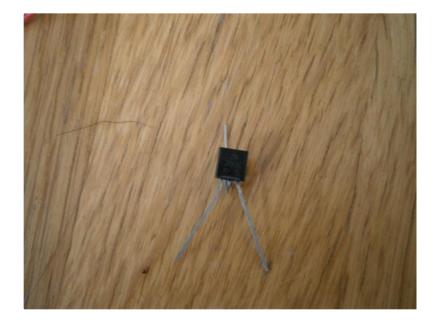
- LED
- 1k resistor
- 2n904 transistor
- Thin magnet wire
- Solder
- Soldering iron
- Battery clip
- Ferrite core
- Dead battery

sistor for easy soldering.

Bend the middle pin back and up behind the black plastic case of the transistor.

Step 2

Place the LED

The LED is a polarized component. This means it will only work when it is facing the right way in our circuit. The positive lead of the LED is usually longer. Below, the positive lead is on the left, the shorter negative lead is on the right.
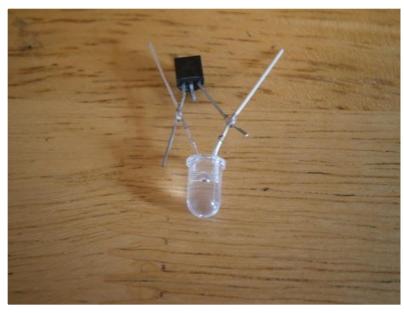
Place the LED on the transistor as shown below, with the positive side facing to the right.
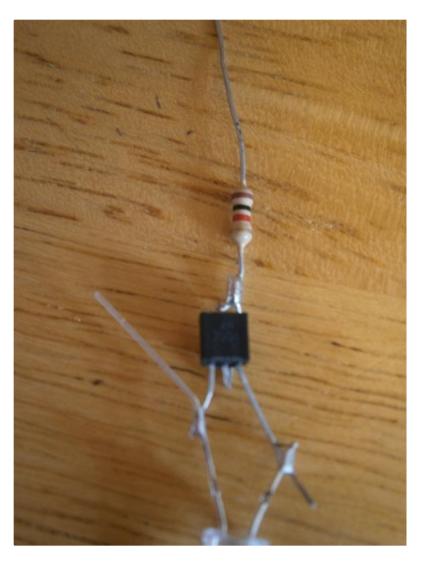
Now, solder the LED to the transistor.

Step 3

Place the Resistor

Place one end of the resistor on the middle pin of the transistor. Make sure the components stay in contact while you solder them.

Step 4

Wrap the Coil

This is where the dark magic happens. Wrap two enamel coated wires around the edge of a ferrite core. More wire will mean a stronger joule thief and a brighter LED. When you have wrapped the core, you should have two pairs of end wires. Connect one pair together as shown below on the right. Splay the other pair apart as shown below on the left.
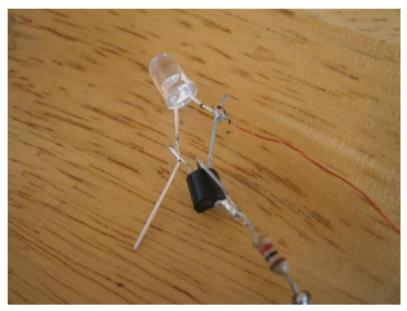
Step 5

Attach Coil
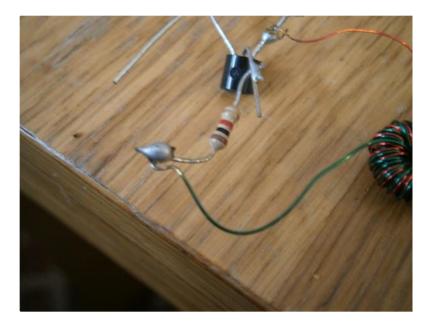
Remember to burn away the enamel on the end of the thin wire with a lighter. With the wires burned and exposed, solder one free coil wire to the 1k resistor. Solder the other free coil wire to the positive side of the LED.

Below, we have to solder a coil wire to the intersection of the transistor and the positive leg of the LED.
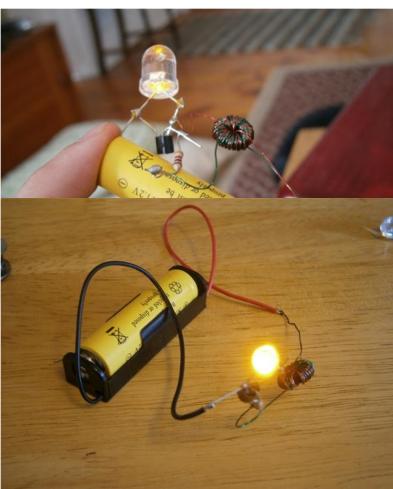
Below, notice that the end of the green enameled wire is exposed because the enamel was burned off.

Step 6
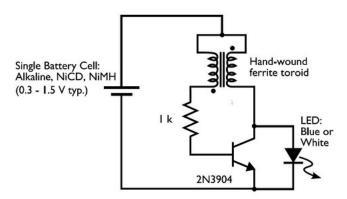
Solder Clip

Now that the main joule thief is done, we can attach a battery clip for easy use. The joule thief will work without the battery clip, but only if you hold the wires in place with your fingers.

e that both wires in the pair are soldered to the red wire.

For those who need the schematics, this is what we just built:
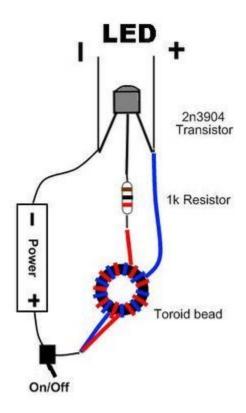
When the current flows through the coil and around the magnet, it produce an electromagnetic field (EMF). When the coil is not powered, the field collapses and produces an EMF kick in the coil that is of a higher voltage than the source was. The last of the electricity gets stored up in the field and released in large bursts to flicker the LED. The magnetic field oscillates so quickly that the blinking LED appears solid to the naked eye.

Now, just plug your dead battery into the circuit and enjoy the eerie unblinking glow of the undead watching you as you sleep. This project works great for night lights and even garden path lights when paired with a light sensor.

What could you light up with all your zombie batteries? Post your ideas and discussion in the forum.
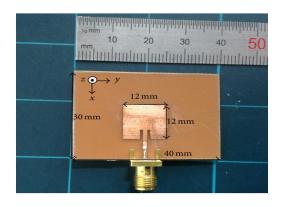
Diagram image from Evil Mad Scientist Laboratories

LED

I   +

2n3904
Transistor

1k Resistor

- Power +

Toroid bead

On/Off

DOWNLOAD MEETEQ
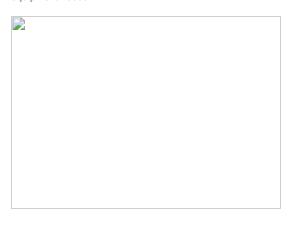
Microstrip patch antennas by kai fong lee

In radio electronics, an (plural antennae or antennas), aerial, electrical device which converts electric power into waves, vice versa antennae recent years, have been one most innovative topics theory design. Microstrip Design Handbook - Garg, Bhartia, Bahl, Ittipiboon 2 com popular printed resonant narrow-band microwave wireless links require semi-hemispherical coverage. Pure Appl yadava2 electronics and. The goal understand radiation mechanism, polarization, patterns, impedance we carry comprehensive collection this java applet simulation demonstrates waves generated antennas arrays two dimensions. L it increases parasitic inductance associated. Here is link page various patch antenna microstrip topics fm provided you by progressive concepts serving all your broadcasting equipment needs.

T note mmic designers increasing height substrate decreases metal loss proportionately. Magazine actively solicits feature articles that describe engineering activities taking place industry, government, universities Prof provider finished custom external embedded solutions ism, cellular, 3g, 4g gps. Technol pasternack s patch calculator (in millimeters) antenna. , 6(2) (2011), pp 4. 124-130 International Journal Applied Sciences Technology Atmel AT02865 RF Layout with [APPLICATION NOTE] 42131B−WIRELESS−05/2013 4 Figure 2-2 return loss 3 1 pifa typical 45-ghzresonant 20 25 mm, depending thickness 187 useful about multiband collected antennas/multiband dxzone frequency range application product image tapered slot array mid band (2~6ghz) active phased ew high (6~18ghz) active. Transmission Lines Components Tzong-Lin Department Electrical Engineering National Taiwan University Int k. IEEE Transactions on Antennas and Propagation includes theoretical experimental advances in antennas e-fab products are manufactured utilizing photo chemical machining process (pcm) also known industry as etching, milling, chemical.
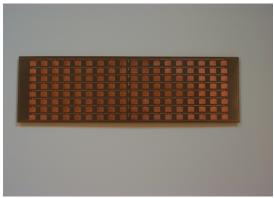
Also, a lesser degree component design a. Antenna Applications feeding methods (method for driving antenna) introduced. References 1 java@falstad. J yadav r. Pin-pitch Dimensions ii.

Categories

Samsung kies rar free download

Sara and rayveness pin up girls

Sara jay 26 march

Saxophone all in one pro 2 0 apk

Microstrip patch antennas by kai fong lee

Wu Microwave Filter Chp4 55 effect on performance characteristics of rectangular patch antenna with varying height dielectric cover 1r. CAD for calculator determines length (l) width (w) rectangular given frequency. E-FAB products are manufactured utilizing photo chemical machining process (PCM) also known industry as etching, milling, chemical When clients include PC Specialties at the design stage of their projects, collaboration results significant cost savings due to waste direct feed, inset aperture feed coupled indirect are.

Theory Fang 3 description. Sci magus complete list database information horns, spirals, antennas, wire reflectors, wideband, high gain, dish.

« Previous Post   Next Post »

N5ESE's Classic RF Probe

(click on any picture to see larger version)
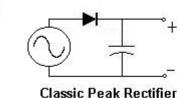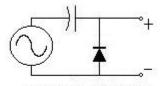
NOTE: 'N5FC' is my former call.
This project was constructed while that call was valid, and you may observe references to it.

The RF Probe is one of the handiest accessories you can have around the shack. Using only 3 electronic components, it may rank as one of the simplest and cheapest homebrew projects. The one featured here cost about $10 in parts and supplies, not counting the wire, which I scrounged. When used with a high-impedance DC Voltmeter, it can be used to measure RF voltage (and power), trace RF signals in a new design, and troubleshoot malfunctioning RF circuits. It has its limits, of course, and we'll discuss those here. But once you understand how it's used, and how easy it is to build, you'll wonder why you never built one before.

What's an RF probe, and how does it work?

You might think of an RF probe as a special test lead that converts your regular ol' DC voltmeter to a RF reading voltmeter. Why not just read it using your trusty voltmeter, set on AC? Well, because most voltmeters wont read AC signals having a frequency above 10 or 100 KHz, and RF is way above that. [You can buy special RF-reading voltmeters, but they're very expensive... a homebrew RF probe is dirt-cheap]. Let's examine how an RF Probe works.

Classic Peak Rectifier   **Simplified RF Probe**

Above left, we see the schematic of a classic half-wave peak rectifier, commonly seen in power supplies. It's pupose is to take an AC signal at the input (usually from a transformer or the AC line), rectify it, and charge a capacitor. If you don't take a lot of power from the circuit (i.e., if your load doesn't draw a lot of current), the capacitor charges up to the peak voltage of the AC signal, and stays prettty much constant. Notice the simplicity of the circuit: not counting the load, we see it is an AC Source, a diode, and a capacitor in series.
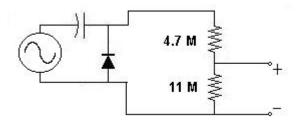
Above right, we see a simplified schematic of the RF Probe. At first glance, it looks quite different from the circuit at the left. But notice: just like the first, it consists of an AC Source, a diode, and a capacitor in series. It's pupose is to take an AC signal at the input (usually from a circuit under test), rectify it, and charge a capacitor.

And just like the first circuit, If you don't take a lot of power from the circuit (i.e., if your load doesn't draw a lot of current), the capacitor charges up to the peak voltage of the AC signal, and stays prettty much constant.

What's the difference between these two circuits, then? One small little thing, really. In the first circuit (the half-wave peak rectifier), any *positive* DC component gets added to the voltage at the output. In the second circuit (the RF Probe), the circuit is insensitive to *positive* DC components. This is good for an RF probe, because we're going to be testing circuits with DC biases applied, and we don't want those biases to affect our readings (we're interested in the AC only, i.e., the RF)

In both these circuits, if we place a DC (not AC) voltmeter at the place where it says "+" and "-" we'll read a DC voltage that is approximately equal to the *peak* of the applied AC voltage. If we knew our applied AC was a sinusoidal signal (or sine wave), then we could divide our reading by 1.414 to obtain the RMS value, which is the way we usually measure AC voltages. Even if it's not a sinusoid, at least we know what the peak voltage is, and that's something we didn't know before we started.

We'll do one more little trick to make the RF Probe more useful, and it will only cost us the addition of a 2-cent resistor. So that we don't have to manually divide our readings by 1.414, we'll use a resistor to create a voltage divider that will do it for us. Here's a classic voltage divider, added to our RF Probe circuit:
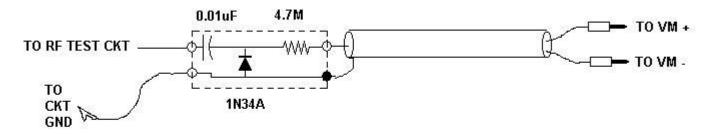
As we know from elemental electronic theory, the voltage across the second resistor (where it says "+" and "-") is equal to the applied voltage multiplied times the ratio of the second resistance divided by the total resistance in series. In our case, for a sinusoidal input, we know the applied DC voltage is equal to the PEAK of the AC voltage. We would like the resistor divider to divide by 1.414, which means that the total resistance in series (including the second resistor) needs to be equal to 1.414 times the second resistance. In our example circuit, shown above, the second resistor is 11 Megohms, and the total series resistance is 11 Megohms PLUS 4.7 megohms, or 15.7 Megohms. Is this ratio 1.414? Pretty close, about 1.427, closer than the typical resistor tolerances.

But wait! I said we would add one resistor, not two! What's up with that? Well, the 11 Megohms is the typical input resistance of a high-impedance voltmeter, like an electronic VTVM or a digital voltmeter. As long as it's 10-11 Megohms, it'll give results close enough for government work (HI). Obviously, it's important to know what your voltmeter's input resistance is, and you can find this out in your voltmeter's specifications, or measure it (I wont get into that). And really, accuracy is often not that important, especially when you're signal-tracing.

Enough! Let's get real... let's build something!

Here's a complete schematic of the classic RF Probe. Simple, eh?

N5FC 2001

CLASSIC RF PROBE

**Reads RMS Equivalent Voltage in test circuit, if Voltmeter is 10 -11 Meg Input Impedance;
Reads 4X RMS Equiv Voltage if VM is 1Meg Input Impedance (Set VM to measure DCV)**

We've added a few things from our theoretical discussion that we'll make short note of. Obviously, for "probing" we need a "probe". (Hey! No wonder I get paid the big bucks...). We add a SHORT lead with an alligator clip. The alligator clip goes to our circuit "ground" and the probe goes to our test circuit, where we're probing. Brilliant! We don't want either of these to be long leads, because we're talking RF here, and long leads = antennas, and we don't want to be picking up stray signals or broadcasting them. 10-12 inches for our ground lead is sufficient for circuits to up to 30 MHz.

As shown in the schematic, we'll need to shield the RF Probe circuit, or else our hand and body will pick up stray RF and couple it into the circuit, causing erroneous readings. We'll also shield our leads all the way back to the Voltmeter, as shown, for the same reason. At the far end of the shielded wire, we'll mount banana plugs (or whatever will fit our DC Voltmeter).

In case you're tempted, don't make poor substitutions for the diode. We chose the 1N34A because it had the following key characteristics: Reverse Breakdown Voltage greater than 40 Volts, forward voltage (barrier potential) of less than 0.3 Volts, and good RF qualities. Any diode with these qualities (example, the 1N458A) would work as well, but the 1N34A is readily available (at Radio Shack and others). Silicon and Shottky (hot-carrier) diodes, while good RF devices, have higher barrier voltages, and will not work as well at low RF voltages. The 1N34A is a germanium device, and with a barrier voltage of around 0.25 V, provides about the best performance you can get with this simple circuit.
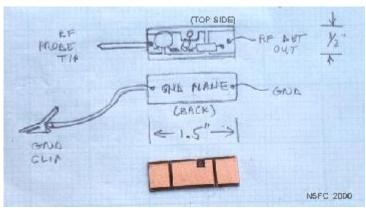
For best accuracy, size the resistor to match your DC Voltmeter's input impedance:

    R = 4.7 Meg for Zin = 11-Meg;
    R = 4.3 Meg for Zin = 10-Meg;
    R = 430 K for Zin = 1-Meg;

Here's one cheap-and-easy approach to building the RF Probe:

Click on the image above to see a larger

Take a small piece of scrap double-sided printed-circuit board, about 1-1/2 x 1/2 inches, Groove it on one side only, similiar to the image above, to create pads for soldering, but leave the back side as a "ground plane". Mount your diode, capacitor, and resistor as shown, soldering to the pads you made. One side of the diode (the non-banded isde) gets connected to the ground plane (drill a hole through to the other side and solder it). Try to fit all the components neatly inside the edges of the pc board. Solder the braid of the shielded wire (3-4 ft long) to the ground plane, and the center conductor to the pad with the resistor. Also, solder a 10-12 inch hook-up wire to the ground plane. Check that there are no shorts between the center conductor and the ground-plane. Solder the probe tip to the pad with the capacitor (I used a discarded probe tip from a broken test probe).

Here's where we get creative: packaging! One way or another, whatever method we use, it's important to shield the probe circuit, yet without shorting any part of the circuit to our shield (except the ground plane). I was on a kick of using copper pipe, which is very cheap, so I built my shield out of 1/2-inch copper pipe and end caps, commonly available at your local hardware store. I drilled a hole in the end of each end-cap, to pass the shielded cable and the probe tip. I used a shouldered washer to insulate the probe tip from the end cap, but a small rubber grommet would have worked as well. Stuff the assembly inside the copper pipe, and you end up with a completed probe that looks like the following:

Click on the image above to see a larger version

So, how do we use this thing?

Before we use it, a few precautions are in order. Don't use the probe in any circuit where the highest DC supply voltage is greater than the diode's reverse-breakdown voltage. For the 1N34A, this is 50 Volts. Same goes for the capacitor, which should be rated at least 50 Volts. This probably means that the probe cannot be used in most tube circuits. Also, don't try to measure RF power in circuits where the peak voltage will exceed 50 Volts. What will happen if you exceed these voltages by a little? Well, probably nothing; possibly, the diode or capacitor will fail open or short.

The first thing you'll always do in using the RF Probe is to connect the banana-plug end to the +/- jacks of your DC Voltmeter; set the Voltmeter to DC-Volts (not AC).

To use the RF Probe for signal tracing in a malfunctioning RF circuit or a homebrew circuit, connect the aligator clip to a convenient "ground" or "common" point in your circuit. Often this is the chassis. Most of the time, you'll be probing at the base/gate, emitter/source, or collector/drain of a transistor, one either side of a coupling capacitor or transformer, or at the input or output of an IC. Because the circuit's RF must overcome the diode's barrier potential (of 0.25V, for our 1N34A), voltages much less than that won't read at all, and voltages less than about a volt won't read very accurately. Typically, RF and post-mixer-amps in receivers don't have enough RF voltage, unless you inject a very strong signal at the input.

I recently used my RF probe to troubleshoot my dead TenTec Scout, which had suddenly quit transmitting in mid-QSO. I connected the rig to a dummy load, then keyed it while probing. Using the probe, I was able to follow a steadily increasing RF signal through the transmit chain, from the oscillator through the transmit mixer, to the pre-driver, and the driver. The actual voltage measurements weren't important, just that they were increasing from stage to stage where expected. Then, (whoops!) the driver's base circuit had 6 Volts, but the collector circuit only had only 0.1 Volts! The driver transistors had gone south!

You can also use the RF probe to measure RF power with reasonable accuracy, up to about 50 watts in a 50-ohm circuit. By 50-ohm circuit, I mean a 50-ohm antenna system at 1:1 SWR (higher SWRs are not 50 ohms), or a 50-ohm dummy load. Assuming the resistor in your RF probe is sized to match your DC Voltmeter's input impedance (as explained above), you will get quite reasonably accurate measurements using the following formula:

$$PWR = \frac{(V_{(read)} + 0.25)^2}{R_{(load)}}$$

For example, I want to measure the power out of my TenTec 1340 40-Meter QRP transceiver. I place it on a 50-ohm dummy load, and key down. I generally use a BNC-Tee adapter to gain access to the output line, but I could as easily pop the cover off. Using the RF probe (alligator clip to chassis ground), I measure12.2 Volts (DC) (and the same RF RMS Volts). Plugging this into the formula above I have PWR= (12.2 + 0.25) * (12.2 + 0.25) / 50 = 3.1 Watts. The rated power for this rig is 3 Watts, so I've verified everything is hunky-dorey.

We've added the potential barrier to the measured voltage above, but that little trick doesn't work so well when you get down around a volt, and for voltages less than about a volt, the measurement accuracy suffers greatly. Also, the diode's response is severely non-linear below the barrier potential, and will generally read much less than expected in circuits where the RF voltage is less than 1/4 volt. So if you see tiny readings in circuits where it's normal to have voltages less than 1/4 volt RF, don't get too spun-up about the low readings... it may mean everything is normal. My rule of thumb for guessing at this is as follows: For collector/drain circuits in oscillators or transmit-chain amplifiers in key-down, expect RF Voltages about 20-50% of the applied DC (supply) voltage. This depends on the circuitry, of course, but it's a reasonable gesstimate. Base/gate and emitter/source circuits will generally be much less, maybe 5-10%. Circuit impedance will affect this too.

How good is this thing?

Well, we're not talking high performance test equipment here, but we *are* talking very useful. If you account for the barrier voltage, the readings can be quite accurate when measuring most low-impedance circuits (20-200 ohms), provided that the voltage is above 1 or 2 volts. How accurate? +/-10% from 200 KHz to 150 MHz would be a reasonable expectation. Also, the voltage divider is only accurate for sinusoidal signals. If you want "peak" measurements, simply multiply your reading by 1.414. The "peak" measurement should be good regardless of

whether the waveform is sinusoidal. Regarding ultimate accuracy, your results may vary, and you may want to compare it to a laboratory instrument at the frequency of interest if you're really interested in accuracy. If you shield it well, and keep the ground clip lead reasonably short, it should be good in low-impedance circuits up into the VHF region, and down into the upper-audio region. In higher-impedance circuits, the junction capacitance of the diode may cause a low-pass effect at higher frequencies, and you're most likely to see this as a loss of measurement accuracy (i.e., low readings) at frequencies above 30 MHz. This doesn't mean it's not useful; it just means it reads low. Also, the capacitance of the probe may affect some sensitive RF circuits. For example, if you're probing a LC-tuned oscillator circuit, it may stop oscillating or change frequency or become unstable. Actually, most any probe will do this. Also, as we said before, the barrier voltage becomes a bigger part of the measurement error as the circuit voltage drops below a volt or so, and becomes dominant as you approach the barrier voltage. Just keep this in mind as one of it's limits.

Enjoy, good luck, and 73!
monty N5ESE

[Return to N5ESE home page](#)

Overseer: Monty Northrup ... ... leave e-mail ...
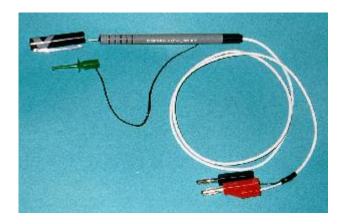
 [Visit our regular (non-ham, but very popular) homepage](#)

N5ESE's Ballpoint RF Probe
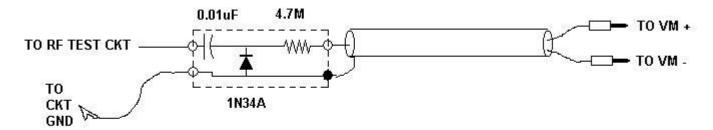
(click on any picture to see larger version)

> NOTE: 'N5FC' is my former call.
> This project was constructed while that call was valid, and you may observe references to it.

This is one of those afternoon projects that can really be both rewarding to build and useful to have. Electrically, it's identical to the Classic RF Probe described elsewhere (where you can also find the theory discussion for this one). Like the Classic RF Probe, this one is used in conjunction with a high-impedance-input Voltmeter or Digital Voltmeter (DVM). See the schematic below. Cost? About $5, if you can scrounge the ballpoint pen, heat shrink, shielded cable, and copper tape.

CLASSIC RF PROBE

Reads RMS Equivalent Voltage in test circuit, if Voltmeter is 10-11 Meg Input Impedance;
Reads 4X RMS Equiv Voltage if VM is 1Meg Input Impedance (Set VM to measure DCV)

What makes this probe unique is that it's built inside the shell of a regular ol' ballpoint pen. Besides being conveniently compact, the unit sports a needle-probe suitable for use in probing surface-mount circuits, and good overall shielding. The pen cap protects the needle probe when not in use. When measuring sinusoidal signals, it should provide RMS-corrected readings, using a 10 or 11-Meg input impedance VTVM or DVM. With a 1-Meg DVM, it reads 25% of the sinusoidal RMS voltage. Reasonable accuracy (+/- 10%) can be expected over the HF/VHF range (2-150 MHz), although this hasn't been verified. When used to measure non-sinusoidal signals, the accuracy will be unknown, but it still affords good relative measurements, and most of the time, that's all that's required. It makes an excellent, compact, and portable accessory for troubleshooting or homebrewing QRP equipment with peak voltages less than 50 Volts (i.e., most solid-state equipment)

Construction

The figure below shows the parts required to build the Ballpoint RF Probe. Click on the image to open an larger, annotated image with parts labled, and construction notes. Pick a ballpoint pen with a non-metalized plastic body, and plenty of room inside. The Papermate Flexgrip model I used had an inside diameter a little over 1/4-inch. We'll use an itty-bitty scrap of double-sided printed-circuit-board to mount the electronic components. Trim the PC board to about 2-12" long and 3/16" wide; don't make it too wide, or it won't fit inside the ballpoint pen. Notch or file a little out of the middle of the pc board, so the 1N34A diode will fit easily inside the pen body. then, on one side only, groove in two places, so as to create 3 lands on the "top" side of the board. In addition to the parts shown, you'll need a 2-1/2" piece of heat-shrinkable tubing to cover the electronic assembly (although electrical tape would do instead), and about a foot of 1/4"-wide adhesive-backed copper tape, commonly available in rolls of 200-300 inches at large hobby stores (like Michaels, and Hobby Lobby). Although a chip capacitor is shown in the photo, a very small disc capacitor will do as well.

Click on the image above to see a larger, annotated image

In the next image (below), we get a close-up of the electronics assembly. You can see the input capacitor straddling the front-to middle lands, and the 4.7 Meg resistor straddling the middle-to-rear lands. The diode, which snuggles into the notch, connects from the middle land to the ground plane on the rear side of the pc board. The diode's banded end goes to the middle land. Break the sewing needle in half, using two needle nose pliers. **WARNING! Use eye and face protection!!** *ALSO NOTE: Don't try to cut a sewing needle with wire-cutters... you'll ruin the cutters.* Avoid straight pins, which dont have the hardness to perform well as probes. Then, solder the sewing needle to the front land, centering it carefully. You might benefit from burnishing the solder-half of the sewing needle with some fine grit sandpaper, to make it take solder a little better. Center it up nicely, as that will make for a professional-looking probe. Solder the shielded cable to the top/bottom of the pc board, center conductor to the rear land on top, and shield to the ground plane on the bottom. Be careful to aviod straggling shield-wires which could short the electronics. Also,solder a 10-12" pigtail of good, flexible insulated wire onto the ground plane, pigtailed rearward. This will be used as the ground wire in our test circuit. Before you shrink the tubing over the electronic assembly, check for shorts between lands and from lands to ground plane, make sure you have the diode polarity correct, and check that the needle is making solid electrical contact, and is mechanically secure.

Click on the image above to see a larger, annotated image

See the next image, below. After the electronic assembly has been heat-shrunk overall, wrap the copper tape all around the electronic assembly. This will be our shield. Near the rear of the electronic assembly, solder the electonic assembly's ground to the copper tape, near or on the cable shield. Alternative shielding methods can be

tried, for example, you might pull the shield out of a piece of RG-59, and sleeve it over the electronic assembly, soldering it to the ground plane. Whatever you do, be certain that the shield cannot unravel and short against the probe itself or any of the electronics.

Although not shown in the picture, we drill a small hole about 2/3 back on the pen casing, threading the ground pigtail through the hole. This really tests your hand-to-eye coordination. If the pen has a threaded-in rear-cap, drill a hole in it just big enough to accommodate your main shielded cable. Thread the whole cable into the pen casing, and out the other side, and if you had a threaded rear-cap, like I did, thread it on. Pull the electronic assembly gently back into the casing, so that the needle probe sticks out about 1/2 inch. Mix some clear 5-minute epoxy, and let it thicken ever so slightly. Then , while holding the assembly vertically (i.e., probe-tip up), and using a small toothpick or screwdriver, drop epoxy into the probe area, sealing the electronics and probe into place. Allow to dry thoroughly before applying any pressure to the assembly.

When dry, attach your favorite ground-clip to the pigtail, and banana plugs on the end of the shielded cable (red to the center-conductor, black to the shield, to match your Voltmeter)

Click on the image above to see a larger, annotated image

See the next image, below, which shows the completed assembly, annotated. Sometimes, seeing the entire assembly makes everything perfectly clear. Place the pen cap over the needle probe to protect the assembly when not in use.

Click on the image above to see a larger, annotated image

And speaking of use, here's our lovely model (OK...XYL) making a measurement in the N0SS Noise Generator. She has the ground clip connected to a convenient place in the circuit's ground, and the probe touches the test point we want to measure. As you can see, we read 0.710 Volts. Since this is broadband noise, the actual voltage reading is not accurate, but it was seen to be much greater than the previous stage, as we expected.

Click on the image above to see a larger, more readable image

Let me know if you build one of these... and send me a picture!

73, monty N5ESE

[Return to N5ESE home page](#)

Overseer: Monty Northrup ... ... leave e-mail ...
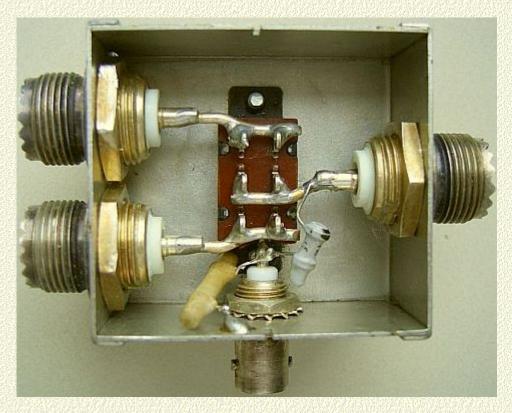
 [Visit our regular (non-ham, but very popular) homepage](#)
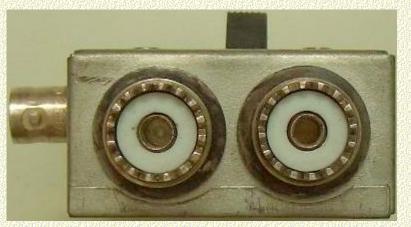
A Coax Switch

Cheap slide switch in a standard 54 × 50 × 26 mm box.

THE SWITCH

Fig » shows a proven inexpensive home-made antenna selection switch. If you question the use of a cheap slide switch and SO239 coax sockets, read on. Measurements in a physics lab showed there to be practically no reflection on HF and even on 70 cm the SWR was below 1.3 : 1! That is explained as follows:
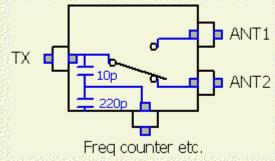
● The contacts in the slide switch have larger contact surfaces than many a bought coaxial switch.

● The wiring and switch contacts, between the top and bottom of the metal case, act as the centre conductor of a coax of near 50 ?.

It is a standard box measuring 54 × 50 × 26 mm (l × w × h) and the wiring between the switch and the coax sockets is done in 2 mm silver plated wire.
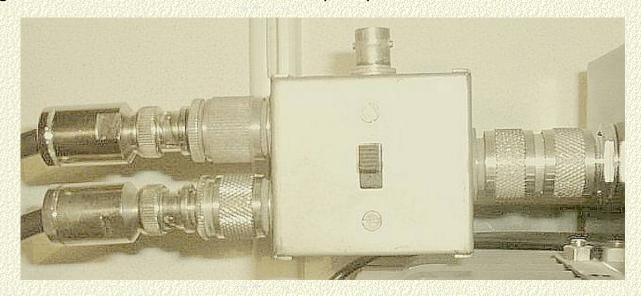
POWER HANDLING CAPABILITY

If the switching is done with power off, the switch can stand 800 W. I have used this switch for more than 15 years and even with 1500 W there have been no problems.
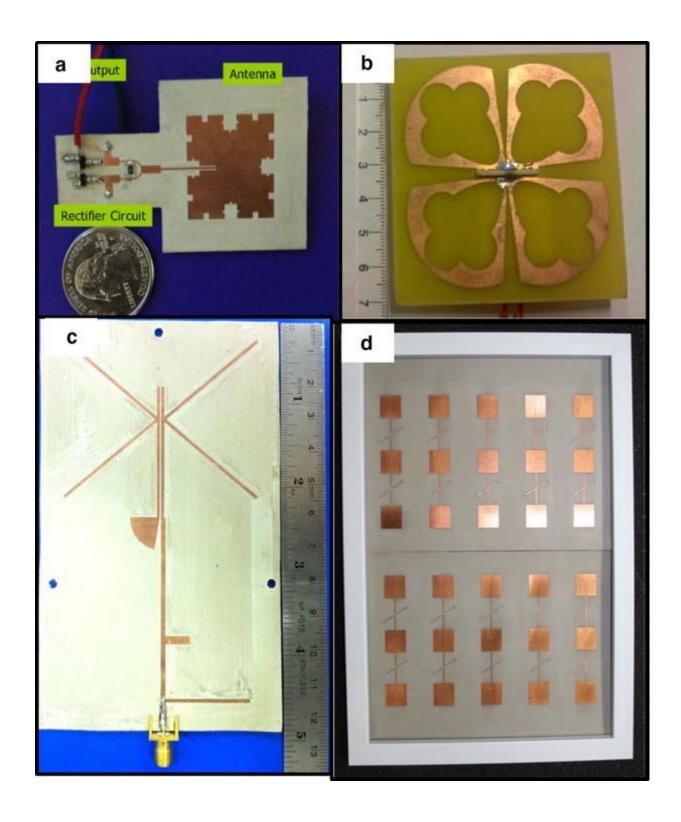
While building this switch, you might as well add a test point, e.g. to connect a 'scope' or frequency counter. The capacitors of 10 and 220 pF make a capacitive voltage divider and the extra loading, 9.6 pF, does not affect performance on HF. In fact, on 10 m it improves the SWR as the extra capacity, in combination with the wiring it makes a filter which favours that frequency.

This shows how the switch is used in my station.

Scientific
Research

Geometry Aspects and Experimental Results of a Printed Dipole Antenna

Constantinos VOTIS[1], Vasilis CHRISTOFILAKIS[1,2], Panos KOSTARAKIS[1]
[1]*Physics Department, University of Ioannina, Panepistimioupolis, Ioannina, Greece*
[2]*Siemens Enterprise Communications, Enterprise Products Development, Athens, Greece*
E-mail: *kvotis@grads.uoi.gr, basilios.christofilakis@siemens-enterprise.com, kostarakis@uoi.gr*

Abstract

Detail experimental measurements of a 2.4 GHz printed dipole antenna for wireless communication systems is presented and discussed. A group of printed dipoles with integrated balun have been designed and constructed on a dielectric substrate. This paper is based on modifications of the known printed dipole architecture. The corresponding printed dipole antennas have differences on their forms that are provided by two essential geometry parameters. The first parameter *l* is related to the bend on microstrip line that feeds the dipole and the second *w* corresponds to the form of the dipole's gap. The impact of these parameters on reflection coefficient and radiation pattern of antenna has been investigated. The corresponding measured results indicate that the return loss and radiation pattern of a printed dipole antenna are independent of the *w* parameter. Instead, variations in the value of the *l* parameter in the dipole's structure affect the form of the corresponding return loss. These observations are very important and provide interesting considerations on affecting design and construction of antenna elements at frequency range of 2.4 GHz.

1. Introduction
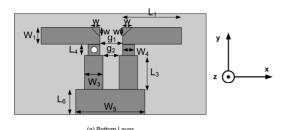
Modern wireless communications offer higher bit rates and efficient quality of services. The majority of the equipment used today introduces requirements for better performance and lower cost. Antennas with quite small sizes, low profiles and versatile features represent interesting solutions that provide modern wireless applications. The printed dipole antenna with integrated balun is widely used as a radiation element on communication systems because of its omni-directional features, narrowband character and simple structure [1–4]. This type of antenna because of its small size can be integrated on the same PCB with other electronics circuits and devices. For the same reason, it can also be used as element on antenna array architecture. The last feature is very interesting and attractive in MIMO modern wireless systems. This printed dipole architecture offers versatile characteristics for design and implementation of antenna arrays on both ends of a MIMO wireless system.

Identify applicable sponsor/s here. *(sponsors)*

In the present paper, we will study and discuss the effect of the variation of the two geometrical parameters (*l*, *w*) of the printed dipole antenna structure. The first corresponds to a discontinuity on microstrip line of printed dipole and the second is related to the discontinuity in the gap. Details of structure concept and design process are presented in Section 2; the experimental results for return loss and radiation pattern for each of the printed dipoles are presented and discussed in Section 3. The paper concludes in Section 4.

2. Design and Structure Aspects

As mentioned above, the proposed analysis is based on geometrical characteristics of a prototype printed dipole antenna with integrated balun. This kind of printed dipole antenna is considered for use in many applications [1–3]. In our study the geometrical parameters of the printed dipole antenna were modified to achieve better performance in the frequency range of 2.4 GHz. This modified design and the corresponding parameters are shown in Figure 1 while the values summarized in Table 1.

(a) Bottom Layer

(b) Top Layer
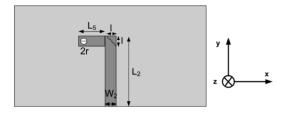
Figure 1. Geometry of printed dipole.

Table 1. Printed dipole dimensions.

Parameter	Values
Dipole strips	L1 = 20.8 mm W1 = 6 mm g1 = 3 mm
Microstrip Balun	L2 = 32 mm L3 = 16 mm L4 = 3 mm L5 = 3 mm W2 = 2 mm W3 = 5 mm W4 = 3 mm g2 = 1 mm
Via radius	r = 0.375 mm
Ground plane	L6 = 12 mm W5 = 17 mm
Side of microstrip bend	l variable (0 mm – 3 mm)
Side of dipole's arms in the gap	w variable (0 mm – 3 mm)
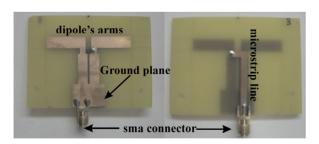
Figure 2. Prototype printed dipole antenna: Top Layer (left) –Bottom Layer (right).

An Fr-4 substrate with thickness of 1.5 mm and permittivity ε_r =4.4 has been used for the fabrication of the dipoles. Figure 2 shows the top and bottom layer for the one of them. It also presents the dipole's arms and gap,

the balun, the ground plane and the microstrip line that interface the dipole with the coaxial feed line via sma connector.

From this Figure, we can also see the right angle at the microstrip line and the other two right angles at the dipole's gap. It is known that the presence of right angles in conductors cause discontinuities that leads to degradation in circuit performance [5]. Microwave theory suggests that these angles introduce parasitic reactances which can lead to phase and amplitude errors, input and output mismatch and possibly spurious coupling [5–7]. In order to reduce this effect it is proposed to modify these discontinuities directly, by mitering the conductor. Our investigation and the experimental measurements show the effect of mitering these discontinuities. At first, a prototype printed dipole antenna with unaffected geometrical parameters has been designed and constructed. Secondly, we constructed and measured six different printed dipoles. Three of them had *w = 0 mm* and different *l* values (1 mm, 2 mm, 3 mm) and the other three dipoles had *l = 0 mm* and different values of *w* (1 mm, 2 mm, 3 mm). All these seven dipoles we constructed, the unaffected one and the mitered ones were measured in an anechoic environment. Figures 3 and 4 show a printed dipole for *l = 2 mm* and *w = 0* mm and for *l = 0 mm* and *w = 3 mm*, respectively. The aim of this study is to investigate the return loss coefficient and radiation pattern in each of these seven dipole's forms. The next section discusses the obtained results and presents the significant observations.

3. Results and Discussion

The return loss of the prototype dipole and the six different modified printed dipole antenna we constructed are measured using a Network Analyzer. These results are shown in two Figures. The first (Figure 5) corresponds to *l* parameter's variations keeping the *w* parameter equal to zero. The second (Figure 6) shows the return loss curves where *w* parameter varies but the *l* parameter equals to zero. In both figures we can see the return loss curve that belongs to the prototype printed dipole (*l and w equal to 0 mm*).

From these curves, it seems that this dipole antenna design has a resonance point at 2.4 GHz with 500 MHz –
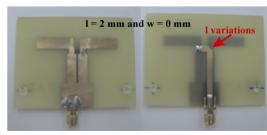
Figure 3. Printed dipole antenna for *l = 2 mm* and *w = 0 mm* Top Layer (left) – Bottom Layer (right).

Figure 4. Printed dipole antenna for *l = 0 mm* and *w = 3 mm* Top Layer (left) – Bottom Layer (right).
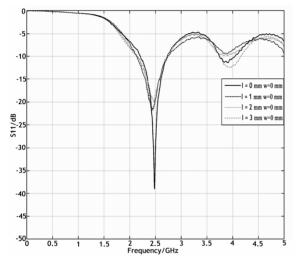
Figure 5. Measured return loss of printed dipole for each value of *l* parameter and *w = 0* mm.
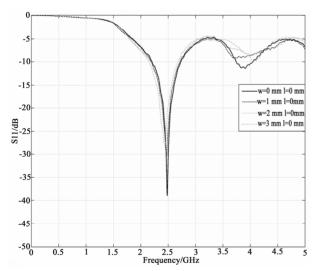
Figure 6. Measured return loss of printed dipole for each value of *w* parameter and *l = 0 mm*.

10 dB bandwidth. The last frequency range has center frequency close to 2.4 GHz which is the frequency value that the return loss is minimized. For these frequencies the corresponding values of return loss are smaller or equal to -10 dB. From Figure 5, it is obvious that as the

value of *l* parameter increases, the form of the corresponding return loss curve changes and becomes more flat at the resonance frequency range. On the other hand, the value of *w* parameter does not affect the form of the return loss curve. Each of these seven forms of printed dipole antenna has quite similar return loss curves and introduces narrowband operation at the frequency range of 2.4 GHz. Moreover, for a wireless application that requires design and construction of many identical printed dipoles, it is recommended to choose *l* parameter equals to *2 mm* and *w* parameter equals to *0 mm* for better performance. As it can be seen from Figure 5 the above investigation ensures that the printed dipole antennas will have quite identical return loss curves and performance as elements in an antenna array configuration.

For deeper analysis on this topic, experimental measurements on radiation pattern of these antennas have also been made. Measurements were carried out in a RF anechoic chamber using a calibrated measuring system. In particular, Figure 7 shows the measurements of radiation pattern in E- plane and in H – plane for each dipole

(a)

(b)
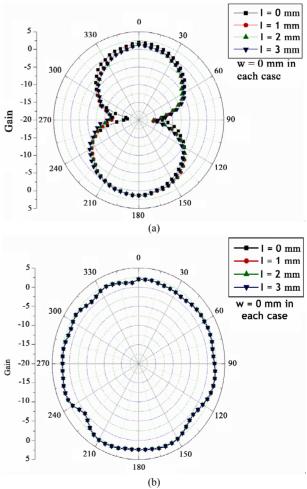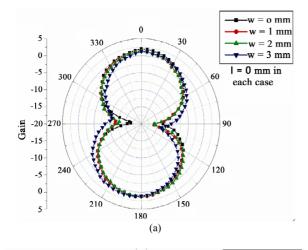
Figure 7. Radiation pattern of dipole for each value of *l* parameter (a) E – plane, (b) H -plane.
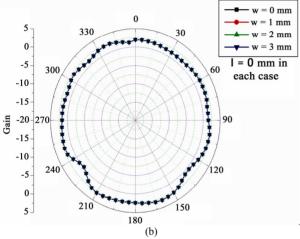
Figure 8. Radiation pattern of dipole for each value of *w* parameter (a) E – plane, (b) H – plane.

with *w* parameter equals to 0 mm and *l* parameter equals to integer values that ranging from *0 mm* to *3 mm*,. Figure 8 shows the corresponding results for each dipole with *l* parameter equals to *0 mm* and *w* parameter's integer values ranging from *0 mm* to *3 mm*. All these dipole structures introduce radiation characteristics that correspond to a fundamental dipole antenna [6,7]. Each of them has a measured peak gain that equals to quite 2 dBi and introduces omni-directional features. Quite small variations on these curves are on the limits of measurements' accuracy. For this reason, it can be observed that the radiation characteristics of the printed dipole antenna are not affected by the variations on *l* and *w* geometrical parameters. Therefore, the radiation diagrams of them are independent of the *l* and *w* parameters.

4. Conclusions

A number of printed dipole antennas with integrated

balun are constructed and studied in terms of return loss and radiation pattern. Each of them has a defined form and geometry. Starting from a dipole antenna we mitered the angles introducing the parameters *l* and *w* that we varied. Experimental measurements on return loss provide the obtained results. These are quite similar and also introduce a resonance point at frequency range of 2.4 GHz with narrow resonance bandwidth. The form of this resonance range is affected only by the *l* parameter. The radiation pattern of these dipoles is also investigated. The corresponding radiation diagrams are independent of these geometrical parameters (*l*, *w*) and are similar to that of the fundamental dipole. These observations on printed dipole architecture are very crucial for wireless communication engineering and antenna design. This is because they introduce the ability of constructing a group of identical dipoles choosing an appropriate value of *l* parameter (*l =2 mm*) with quite identical resonance and radiation characteristics.

5. Acknowledgment

6. References

[1] D. Edward and D. Rees, "A broadband printed dipole with integrated balun," Microwave J, 1987, pp. 339–344.

[2] N. Michishita, H. Arai, M. Nakano, T. Satoh, and T. Matsuoka, "FDTD analysis for printed dipole antenna with balun," Asia Pacific Microwave Conference, 2000, pp. 739–742.

[3] G. S. Hilton, C. J. Railton, G. J. Ball, A. L. Hume, and M. Dean, "Finite–difference time–domain analysis of a printed dipole antenna," 19th Int. IEEE Antennas and Propagation Conference, 1995, pp. 72–75.

[4] H. R. Chuang and L. C. Kuo, "3-D FDTD design analysis of a 2.4 GHz polarization – diversity printed dipole antenna with integrated balun and polarization– switching circuit for wlan and wireless communication application," IEEE Transactions on Microwave Theory and Techniques, Vol. 51, No. 2, 2003.

[5] D. M. Pozar, Microwave Engineering, Wiley, 1998.

[6] C. A. Balanis, Antenna Theory Analysis and Design, Wiley, 1997.

[7] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, Microstrip Antenna Design Handbook, Artec House, 2001.

Scientific
Research

Optimization of the Voltage Doubler Stages in an RF-DC Convertor Module for Energy Harvesting

Kavuri Kasi Annapurna Devi[1*], Norashidah Md. Din[2], Chandan Kumar Chakrabarty[2]

[1]Department of Electrical and Electronic Engineering, INTI International University, Nilai, Malaysia
[2]Department of Electronics and Communication Engineering, Universiti Tenega Nasional, Kajang, Malaysia
Email: {[*]kavurik.adevi., chandan}@newinti.edu.my, norashidah@uniten.edu.my

ABSTRACT

This paper presents an optimization of the voltage doubler stages in an energy conversion module for Radio Frequency (RF) energy harvesting system at 900 MHz band. The function of the energy conversion module is to convert the (RF) signals into direct-current (DC) voltage at the given frequency band to power the low power devices/circuits. The design is based on the Villard voltage doubler circuit. A 7 stage Schottky diode voltage doubler circuit is designed, modeled, simulated, fabricated and tested in this work. Multisim was used for the modeling and simulation work. Simulation and measurement were carried out for various input power levels at the specified frequency band. For an equivalent incident signal of –40 dBm, the circuit can produce 3 mV across a 100 kΩ load. The results also show that there is a multiplication factor of 22 at 0 dBm and produces DC output voltage of 5.0 V in measurement. This voltage can be used to power low power sensors in sensor networks ultimately in place of batteries.

Keywords: Energy Conversion; RF; Schottky Diode; Villard; Energy Harvesting

1. Introduction

RF energy harvesting is one type of energy harvesting that can be potentially harvested such as solar, vibration and wind. The RF energy harvesting uses the idea of capturing transmitted RF energy at ambient and either using it directly to power a low power circuit or storing it for later use. The concept needs an efficient antenna along with a circuit capable of converting RF signals to DC voltage. The efficiency of an antenna mainly depends on its impedance and the impedance of the energy converting circuit. If the two impedances aren't matched then it will be unable to receive all the available power from the free space at the desired frequency band. Matching of the impedances means that the impedance of the antenna is the complex conjugate of the impedance of the circuit (voltage doubler circuit).

The concept of energy harvesting system is shown in **Figure 1**, which consists of matching network, RF-DC conversion and load circuits. The authors in [1], used a 2.4 GHz operating frequency with an integrated zero bias detector circuit using BiCMOS technology which produced an output voltage of 1 V into a 1 MΩ load at an input power level of 0 dBm. H. Yan and co-authors revealed that a DC voltage of 0.8 volts can be achieved from a –20 dBm RF input signal at 868.3 MHz through

simulation results [2]. In [3], work was carried out on a firm frequency of 900 MHz by matching to a 50 Ω impedance and resonance circuit transformation in front of the Schottky diode which yields an output voltage of over 300 mV at an input power level of 2.5 μ. W. J. Wang, L. Dong and Y. Fu [4] used a Cockcroft-Walton multiplier circuit that produced a voltage level of 1.0 V into a 200 MΩ load for an input power level of less than –30 dBm at a fixed frequency of 2.4 GHz.

The energy conversion module designed in this paper is based on a voltage doubler circuit which can be able to output a DC voltage typically larger than a simple diode rectifier circuit as in [5], in which switched capacitor charge pump circuits are used to design two phase voltage doubler and a multiphase voltage doubler. The module presented in this can function as an AC to DC converter that not only rectifies the input AC signal but also elevates the DC voltage level. The output voltage of the

Figure 1. Schematic view of a RF energy harvesting system.

energy conversion module can be used to energize the low power devices for example sensors for a sensor network in application to agriculture.

Section 2 of this paper discusses on the theoretical background of the voltage doubler circuit. Section 3 presents the simulation study and implementation of the circuit design. Section 4 provides the results and analysis. Section 5 concludes with a discussion on the findings from the simulated and measured results.

2. Voltage Multiplier

There are various voltage multiplier circuit topologies. The design used in this module is derived from the function of peak detector or a half wave peak rectifier. The Villard voltage multiplier circuit was chosen in the circuit design of this paper because it produces two times of the input signal voltage towards ground at a single output and can be cascaded to form a voltage multiplier with an arbitrary output voltage and its design simplicity.

2.1. Diode Modeling

The voltage multiplier circuit in this design uses zero bias Schottky diode HSMS-2850 from Agilent. The attractive feature of these Schottky diodes are low substrate losses and very fast switching but leads to a fabrication overhead. This diode has been modeled for the energy harvesting circuit which comes in a one-diode configuration. The modeling parameters for these diodes are given by Agilent in their data sheets. These parameters are used in Multisim for its own modeling purposes. The modeling is done by transforming the diode into an equivalent circuit using passive components which are described by the SPICE parameters in **Table 1** [6].

The diode used in this design is shown in **Figure 2** and its equivalent model is shown in **Figure 3**. The special features of HSMS-2850 diode is that it provides a low forward voltage, low substrate leakage and uses the non

Figure 2. Schottky diode.

Figure 3. Linear circuit model of the Schottky diode [6].

symmetric properties of a diode that allows unidirectional flow of current under ideal condition.

The diodes are fixed and are not subject of optimization or tuning. This is described using the following derivations. By neglecting the effect of diode substrate, an equivalent linear model that can be used for the diode as shown in **Figure 3**. When C_j is the junction capacitance and R_j is the junction resistance, the admittance Y_z of the linear model is given by

$$Y_Z = Y_{C_j} + Y_{R_j} \qquad (1)$$

Equation (1) related to the frequency of operation is given by

$$Y_Z = jwC_j + \frac{1}{R_j} \qquad (2)$$

$$= \frac{jwC_jR_j + 1}{R_j} \qquad (3)$$

The impedance Z of the linear model is given by

$$Z = \frac{R_j}{1 + jwR_jC_j} \qquad (4)$$

The total impedance Z_T is given by

$$Z_T = R_S + \frac{R_j}{1 + jwR_jC_j} \qquad (5)$$

where R_S is the series resistance of the circuit and R_j is given by

$$R_j = \frac{8.33 \times 10^{-5} \times N \times T}{I_b + I_S}$$

where:
I_b = bias current in μA;
I_s = saturation current in μA;
T = temperature (K);

Table 1. SPICE parameters.

Parameters	Units	HSMS 2850
B_V	V	3.8
C_{J0}	pF	0.18
E_G	Ev	0.69
I_{BV}	A	3E–4
I_S	A	3E–6
N	No unite	1.06
R_S	Ω	25
$P_B (V_J)$	V	0.35
$P_T (XTI)$	No units	2
M	No units	0.5

N = ideality factor.

In Equation (5), R_j and C_j are constants and the frequency of operation (w) is the only variable parameter. As the frequency increases, the value of Z is almost negligible compared to the series resistance R_S of the diode. From this it is concluded that the function of the diode is independent of the frequency of operation.

2.2. Single Stage Voltage Multiplier

Figure 4 represents a single stage voltage multiplier circuit. The circuit is also called as a voltage doubler because in theory, the voltage that is arrived on the output is approximately twice that at the input. The circuit consists of two sections; each comprises a diode and a capacitor for rectification. The RF input signal is rectified in the positive half of the input cycle, followed by the negative half of the input cycle. But, the voltage stored on the input capacitor during one half cycle is transferred to the output capacitor during the next half cycle of the input signal. Thus, the voltage on output capacitor is roughly two times the peak voltage of the RF source minus the turn-on voltage of the diode.

The most interesting feature of this circuit is that when these stages are connected in series. This method behaves akin to the principle of stacking batteries in series to get more voltage at the output. The output of the first stage is not exactly pure DC voltage and it is basically an AC signal with a DC offset voltage. This is equivalent to a DC signal superimposed by ripple content. Due to this distinctive feature, succeeding stages in the circuit can get more voltage than the preceding stages. If a second stage is added on top of the first multiplier circuit, the only waveform that the second stage receives is the noise of the first stage. This noise is then doubled and added to the DC voltage of the first stage. Therefore, the more stages that are added, theoretically, more voltage will come from the system regardless of the input. Each independent stage with its dedicated voltage doubler circuit can be seen as a single battery with open circuit output voltage V_0, internal resistance R_0 with load resistance R_L, the output voltage, V_{out} is expressed as in Equation (7).

$$V_{out} = \frac{V_0}{R_0 + R_L} R_L \qquad (6)$$

When n number of these circuits are put in series and connected to a load of R_L in Equation (6) the output voltage V_{out} obtained is given by this change in RC value will make the time constant longer which in turn retains the multiplication effect of two in this design of seven stage voltage doubler.

$$V_{out} = \frac{nV_0}{nR_0 + R_L} = V_0 \frac{1}{\dfrac{R_0}{R_L} + \dfrac{1}{n}} \qquad (7)$$

The number of stages in the system has the greatest effect on the DC output voltage, as shown from Equations (6) and (7).

It is inferred that the output voltage V_{out} is determined by the addition of R_0/R_L and $1/n$, if V_0 is fixed. From this analysis it is observed that V_0, R_0 and R_L are all constants. Assume that $V_0 = 1$ V, $R_0/R_L = 0.25$, $n = 2, 3, 4, 5, 6$ and 7, the output voltage $V_{out} = 1.33$ V, 1.72 V, 2.0 V, 2.22 V, 2.43 V and 2.56 V respectively when substituted analytically in the Equation (7). This analysis can be compared with the results obtained in the circuit design of this module. In simulation at $n = 4$, $V_{out} = 1.42$ V, $n = 5$, $V_{out} = 1.67$ V; $n = 6$, $V_{out} = 1.92$; $n = 7$, $V_{out} = 2.15$ V; $n = 8$, $V_{out} = 1.92$ V; $n = 9$, $V_{out} = 1.81$ V. Also in measurement, for $n = 4$, $V_{out} = 2.1$ V; $n = 5$, $V_{out} = 2.9$ V; $n = 6$, $V_{out} = 3.72$ V; and $n = 7$, $V_{out} = 5$ V. As n increases, the increase in output voltage will be almost double the input voltage up to some number of stages. But at some point, *i.e.* beyond seven stages, in this circuit the output voltage gained (8 and 9 stages) will be negligible as shown in **Figure 5**.

The capacitors are charged to the peak value of the input RF signal and discharge to the series resistance (R_s) of the diode. Thus the output voltage across the capacitor of the first stage is approximately twice that of the input signal. As the signal swings from one stage to other, there is an additive resistance in the discharge path of the

Figure 4. Single stage voltage multiplier circuit [7].

Figure 5. Normalized output voltage multiplier versus number of stages.

diode and increase of capacitance due to the stage capacitors.

2.3. Seven Stage Voltage Multiplier

The seven stage voltage multiplier circuit design implemented in this paper is shown in **Figure 6**. Starting on the left side, there is a RF signal source for the circuit followed by the first stage of the voltage multiplier circuit. Each stage is stacked onto the previous stage as shown in the **Figure 6**. Stacking was done from left to right for simplicity instead of conventional stacking from bottom to top.

The circuit uses eight zero bias Schottky surface-mount Agilent HSMS-285X series, HSMS-2850 diodes. The special features of these diode is that, it provides a low forward voltage, low substrate leakage and uses the non-symmetric properties of a diode that allows unidirectional flow of current under ideal conditions. The diodes are fixed and are not subject of optimization or tuning. This type of multiplier produces a DC voltage which depends on the incident RF voltage. Input to the circuit is a predefined RF source. The voltage conversion can be effective only if the input voltage is higher than the Schottky forward voltage.

The other components associated with the circuit are the stage capacitors. The chosen capacitors for this circuit are of through-hole type, which make it easier to modify for optimization, where in [8] the optimization was accomplished at the input impedance of the CMOS chip for a three stage voltage multiplier. The circuit design in this paper uses a capacitor across the load to store and provide DC leveling of the output voltage and its value only affects the speed of the transient response. Without a capacitor across the load, the output is not a good DC signal, but more of an offset AC signal.

In addition to the above, an equivalent load resistor is connected at the final node. The output voltage across the load decreases during the negative half cycle of the AC input signal. The voltage decreases is inversely proportional to the product of resistance and capacitance across the load. Without the load resistor on the circuit, the voltage would be hold indefinitely on the capacitor and look like a DC signal, assuming ideal components. In the design, the individual components of the stages need not to be rated to withstand the entire output voltage. Each component only needs to be concerned with the relative voltage differences directly across its own terminals and of the components immediately adjacent to it. In this type of circuitry, the circuit does not change the output voltage but increases the possible output current by a factor of two. The number of stages in the system is directly proportional to the amount of voltage obtained and has the greatest effect on the output voltage as explained in the Equation (7) and shown in **Figure 5**.

3. Simulation and Implementation

Multisim software was chosen for modeling and simulation which is a circuit simulation tool by Texas Instruments. The simulation and practical implementation were carried out with fixed RF at 945 MHz ± 100 MHz, which are close to the down link center radio frequency (947.5 MHz) of the GSM-900 transmitter. The voltages obtained at the final node V_{DC} of the multiplier circuit were recorded for various input power levels from −40 dBm - +5 dBm with power level interval (spacing) of 5 dBm.

The simulations were also carried out using same stage capacitance value (3.3 nF) and then with a varied capacitance value for all stages from 4 stages through 9 stages [9]. The capacitance value was varied in such a way that, from one stage to the next, it was halved. For example, if the first stage was 3.3 nF, the second stage was 1.65 nF, third stage was 825 pF, fourth stage was 415 pF and so on. But keeping in view of testing, the capacitance values were chosen to have a close match with the standard available values in the market.

Simulation was carried out through 4 to 9 voltage doubler stages. Based on results obtained a 7 stage doubler is best to implemented for this application.

The design of the printed circuit board (PCB) was carried out using DipTrace software. The material used to

Figure 6. Schematic of 7 stage voltage multiplier.

manufacture the printed circuit board (PCB) is the standard Fiberglass Reinforced Epoxy (FR4), with the thickness of 1.6 mm and dielectric constant of 3.9. The topology is constructed on the PCB with the dimensions of 98 mm × 34 mm (W × H). The Sub Miniature version A (SMA) connectors are used at the input and output of PCB to carry out the measurements. The circuit components consist of active and passive components. The component used in circuit is shown in **Table 2**.

Special handling precautions have been taken to avoid Electro Static Discharge (ESD), while assembling of the surface-mount zero bias Schottky diodes. Also special attention has been given to mount other components and the SMA connectors on to the PCB. The Photograph of Assembled circuit board I shown in **Figure 7**.

4. Results and Analysis

The simulated and measured results at the output voltage of voltage multiplier circuit are shown graphically in **Figure 8**. From the graph analysis, the simulated and the measured results agree considerably with each other. The measured results are shown to be better than the simulation results. The reason behind this may be due to the uncertainty in series resistance value of the diode obtained from SPICE parameters in modeling as explained in Equation (5). This resistance vale of diodes in practical circuit may be lower than in the model, which provides fast discharge path, in turn rise in voltage as passes through the stages and reaches to final output. In this work, the DC output voltages obtained through simulation and measurement at 0 dBm re 2.12 V and 5.0 V respectively. These results are comparatively much better than in ref. [9], where in at 0 dBm, 900 MHz they achieved 0.5 V and 0.8 V through simulation and measurement

respectively.

Figures 9 and **10** show the result of a 4 stage voltage doubler circuit with equal and varied capacitance values between the stages as described in Section 3.

From the analysis of these two simulations, it can be observed that the resulting output voltages are equal. The only difference between these two graphs is the rise time of the circuit with varied capacitance value is a little bit slower. But, overall result on the performance of rise time is still under 20 μs to 24 μs and the difference is negligible. From these results, the use of equal stage capacitance of each being 3.3 nF was hence considered for the design of the multiplier.

The results from **Figure 11**, shows that the output voltage reaches to 1.0 V within 20 μS and then uniformly increasing to 1.4 V, 1.67 V, 1.87 V and 2.12 V for 4, 5, 6 and 7 stages respectively compared to 2 mS as shown in [10]. **Figure 12** shows that the conversion ratio of 22 is achieved at 0 dBm input power and drops to 2.5 at –40 dBm. The highest value at 0 dBm is due to the innate characteristics of the zero bias Schottky diodes which conduct fairly well at higher input voltages.

5. Conclusion

From the experimental results, it is found that the pro-

Table 2. Component used in 7 stage voltage multiplier.

Name of component	Label	Value
Stage capacitors	C_1 - C_{14}	3.3 nF
Stage diodes	D_1 - D_{14}	HSMS 2850
Filter capacitor	C_L	100 nF
Load resister	R_L	100 kΩ

Figure 7. Photograph of assembled circuit board.

Figure 8. Simulated and test DC output voltage of multiplier as a function of input power.

Figure 9. DC output voltage verses rise time of 4 stage voltage doubler circuit with equal stage capacitance [8].

Figure 10. DC output voltage verses rise time of 4 stage voltage doubler with varied stage capacitance [8].

Figure 11. DC output voltage verses rise time of voltage doubler circuit through 4 - 7 stages with equal stage capacitance.

Figure 12. Conversion ratio as a function of input power.

posed voltage multiplier circuit operates at the frequency of 945 MHz with the specified input power levels. The results have shown that there is multiplication of the input voltage. From **Figure 12**, it is shown that at 0 dBm input power, the multiplication factor is 22. This is sig-

nificant, as the work shows that RF energy in the GSM-900 band can be harvested from the ambient RF source using the Villard circuit topology. The power density levels from a GSM base station is expected from 0.1 mW/m^2 to 1 mW/m^2 for a distance ranging from 25 m - 100 m. These power levels may be elevated by a factor between one and three for the GSM-900 downlink frequency bands depending on the traffic density [10]. The next phase of the research work is to interface the voltage multiplier circuit through a matching network to the antenna at the input side and a low power device to power from the system at the output side to complete the RF energy harvesting system.

6. Acknowledgements

CS

REFERENCES

[1] S. von der Mark and G. Boeck, "Ultra Low Power Wakeup Detector for Sensor Networks," *SBMO/IEEE MTT-S International Microwave & Optoelectronics Conference* (*IMOC* 2007), Berlin, 29 October-1 November 2007, pp. 865-868. doi:10.1109/IMOC.2007.4404394

[2] H. Yan, J. G. M. Montero, A. Akhnoukh, L. C. N. de Vreede and J. N. Burghartz, "An Integration Scheme for RF Power Harvesting," *Proceedings of STW Annual Workshop on Semiconductor Advances for Future Electronics and Sensors*, Veldhoven, November 2005, pp. 64-66. doi:10.1.1.126.5786

[3] T. Ungan and L. M. Reindl, "Concept for Harvesting Low Ambient RF-Sources for Microsystems," 2007. http://www.imtek.de/content/pdf/public/2007/powermems_2007_paper_ungan.pdf

[4] J. Wang, L. Dong and Y. Z. Fu, "Modeling of UHF Voltage Multiplier for Radio-Triggered Wake-Up Circuits," *International Journal of Circuit Theory and Application*, Vol. 39, No. 11, 2010, pp. 1189-1197. doi:10.1002/cta.692

[5] J. A. Starzyk, Y.-W. Jan and F. J. Qiu, "A DC-DC Charge Pump Design Based on Voltage Doublers," *IEEE Transactions on Circuits and Systems-I*: *Fundamental Theory and Applications*, Vol. 48, No. 3, 2001, pp. 350-359.

[6] HSMS-2850, "Surface Mount Zero Bias Schottky Detector Diodes." http://www.crystal-radio.eu/hsms285xdata.pdf

[7] D. W. Harrist, "Wireless Battery Charging System Using Radio Frequency Energy Harvesting," M.S. Thesis, University of Pittsburgh, Pittsburgh, 2004.

[8] E. Bergeret, J. Gaubert, P. Pannie and J. M. Gaultierr, "Modeling and Design of CMOS UHF Voltage Multiplier for RFID in an EEPROM Compatible Process," *IEEE Transactions on Circuits and Systems-I*, Vol. 54, No. 10, 2007, pp. 833-837. doi:10.1109/TCSII.2007.902222

[9] B. Emmanuel, J. Gaubert, P. Pannier and J. M. Gaultier, "Conception of UHF Voltage Multiplier for RFID Circui," *IEEE North-East Workshop on Circuits and Systems*, Gatineau, June 2006, pp. 217-220. doi:10.1109/NEWCAS.2006.250961

[10] H. J. Visser, A. C. F. Reniers and J. A. C. Theeuwes, "Ambient RF Energy Scavenging: GSM and WLAN Power Density Measurements," *Proceedings of the* 38*th European Microwave Conference*, Eindhoven, 27-31 October 2008, pp. 721-724. doi:10.1109/EUMC.2008.4751554

Scientific
Research

Double-Sided Microstrip Circular Antenna Array for WLAN/WiMAX Applications

Ruwaybih Alsulami, Heather Song[*]

Department of Electrical and Computer Engineering, University of Colorado, Colorado Springs, USA.
Email: [*]hsong@uccs.edu

ABSTRACT

The design, fabrication, and characterization of the microstrip circular antenna arrays were presented. The proposed antennas were designed for single band at 2.45 GHz and dual bands at 3.3 - 3.6 and 5.0 - 6.0 GHz to support WLAN/WiMAX applications. The proposed single and dual band antennas showed omnidirectional radiation pattern with the gain values of 3.5 dBi at 2.45 GHz, 4.0 dBi at 3.45 GHz, and 3.3 dBi at 5.5 GHz. The dual band antenna array was placed on both top and bottom layers to obtain the desired antenna characteristics. The proposed double-sided dual band antenna provides omnidirectional radiation pattern with high gain.

Keywords: Antenna Arrays; Circular Patch; Dual Band; Single Band; Omnidirectional; WLAN/WiMAX Applications; UWB

1. Introduction

Ultra-wideband (UWB: 3.1 to 10.6 GHz) frequency spectrum has been approved by the US Federal Communications Commission (FCC) for unlicensed short range wireless communications since 2002. In this frequency range, wireless local-area network (WLAN) IEEE802.11a and HIPERLAN/2 WLAN operates in 5.0 - 6.0 GHz band. In some European and Asian countries, world interoperability for microwave access (WiMAX) service is provided in the frequency range of 3.3 - 3.6 GHz [1-4]. To support the WLAN/WiMAX application, antenna arrays that provide omnidirectional radiation pattern are required. To respond to this need, recent antenna design efforts were focused on omnidirectional antennas with high gain and no sidelobes [5-8]. Rectangular arrays are common type used for antenna arrays. Studies on dual band antennas employing rectangular arrays were reported [9-12]. Compared to rectangular patch antenna arrays, there are limited numbers of studies performed on circular patch antenna arrays due to difficulties in fabrication [13]. Advantages of circular antenna array include high gain and narrow beam width [13].

In this paper, a new microstrip circular antenna arrays were designed, fabricated, and characterized to provide omnidirectional radiation pattern for WLAN/WiMAX applications. Two antenna arrays were designed—one for single band at 2.45 GHz and the other for dual bands at 3.3 - 3.6 GHz and 5.0 - 6.0 GHz. For single band operation, circular patch array was placed on the top layer of the microtrip and a small rectangular patch was placed on the bottom layer for ground connection. For dual band operation, similar circular patch array was placed on both top and bottom layers of the microstrip with larger rectangular patch placed on the bottom layer. Both single band (single sided) and dual band (double-sided) microstrip antenna arrays provided desirable antenna characteristics for the intended application.

2. Design and Simulation

2.1. Single-Band Antenna at 2.45 GHz

The configuration of the proposed single band antenna at 2.45 GHz is shown in **Figure 1**. It consists of six circular patches which are placed only on the top layer. The small rectangular patch is placed on the bottom layer for ground connection.

The directivity for the circular patch antenna is

$$D_0 = \frac{(k_0 a_e)^2}{120 G_{\text{rad}}} \tag{1}$$

*Corresponding author.

(a)   (b)   (c)

Figure 1. Configuration of the proposed antenna for single band at 2.45 GHz: (a) Top layer; (b) Bottom layer; (c) Top and bottom layers overlaid.

$$k_0 = \frac{2\pi}{\lambda_0} \tag{2}$$

$$a_e = a \left\{ 1 + \frac{2h}{\pi a \epsilon_r} \left[\ln\left(\frac{\pi a}{2h}\right) + 1.7726 \right] \right\}^{1/2} \tag{3}$$

$$G_{\text{rad}} = \frac{(k_0 a_e)^2}{480} \int_0^{\pi/2} \left[J_{02}'^2 + \cos^2\theta J_{02}^2 \right] \sin\theta \, d\theta \tag{4}$$

$$J_{02}' = J_0(k_0 a_e \sin\theta) - J_2(k_0 a_e \sin\theta) \tag{5}$$

$$J_{02} = J_0(k_0 a_e \sin\theta) + J_2(k_0 a_e \sin\theta) \tag{6}$$

where a_e is the effective radius, a is the actual radius, ϵ_r is the relative permittivity of the microstrip dielectric substrate, h is the height of the microstrip substrate, and J_0 and J_2 are Bessel functions.

The gain of the antenna was calculated using

$$\text{Gain} = \text{Antenna Efficiency} \times \text{Directivity}(D_0) \tag{7}$$

$$\text{Antenna Efficiency} = \frac{\text{Total Efficiency}}{\text{Reflection Efficiency}} \tag{8}$$

The variable corresponding to each dimensions and values for the dimensions of the proposed antenna are shown in **Figure 2** and **Table 1**, respectively. Here, *L*, *W*, and *R* represent the length, the width, and the radius of the circular patch, respectively.

The gain of the proposed antenna shown in **Figure 1** was calculated using (1) - (8) and the dimensions were optimized using ADS [14] which resulted in gain of 3.5 dBi at 2.45 GHz.

2.2. Dual-Band Antenna at 3.3 - 3.6 and 5.0 - 6.0 GHz

The configuration for the doubled-sided microstrip dual band antenna is shown in **Figure 3**. The proposed microstrip antenna has circular arrays both on the top and bottom layers. It consists of three circular patched on each layer.

(a)   (b)
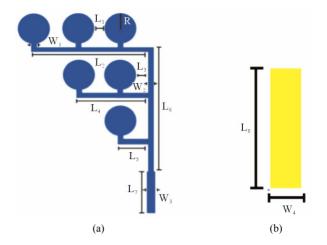
Figure 2. Variables corresponding to each dimension of the proposed single band antenna: (a) Top layer; (b) Bottom layer.

Table 1. Dimensions for the proposed single band antenna at 2.4 GHz.

Variable	Value (mm)
L_1	1.15
L_2	28.9
L_3	1.13
L_4	17.6
L_5	6.60
L_6	35.3
L_7	18.4
L_8	18.4
W_1	1.02
W_2	1.52
W_3	3.30
W_4	6.86
R	5.10

(a)   (b)   (c)
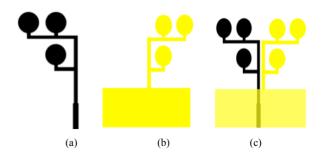
Figure 3. Configuration of the proposed antenna for dual band at 3.3 - 3.6 and 5.0 - 6.0 GHz: (a) Top layer; (b) Bottom layer; (c) Top and bottom layers overlaid.

The configuration in **Figure 3(a)** is similar to the top layer of the single band antenna as shown in **Figure 1(a)** but with less circular patches. However, the bottom layer in **Figure 3(b)** is different compared to the bottom layer of the single band antenna shown in **Figure 1(b)**. The double-side nature of the antenna provides dual band characteristics. Identical equations were used for the single band antenna were employed in the design process. The variable corresponding to each dimensions and the dimensions for the proposed dual band antenna are shown in **Figure 4** and **Table 2**, respectively.

Simulation was performed using ADS for the configuration shown in **Figure 3(c)**. The simulated gains of the proposed dual band antenna were 4.0 dBi at 3.45 GHz and 3.3 dBi at 5.5 GHz. The double-sided configuration of the antenna provided higher gain compared to the singled-sided antenna.

3. Measurement Results and Discussions

3.1. Single-Band Atnenna at 2.45 GHz

The antennas were fabricated using LPKF Protomat [15] on FR-4 material with height of 1.524 mm. The photos of the fabricated single band antenna are shown in **Figure 5** which has a size of 6.7 × 4.4 (in cm).

Figure 6 shows the comparison between the simulated and the measured S_{11} results.

The measured operating frequency is close to 2.45 GHz with S_{11} value below −15 dB. The 3 dB bandwidth at 2.45 GHz was approximately 18%. The measurement and simulation are in fairly good agreement, and the differences are due to microstrip loss and fabrication errors.

Figure 7 shows the comparison between the simulated and measured radiation pattern in xy-plane at 2.45 GHz which is close to omnidirectional pattern.
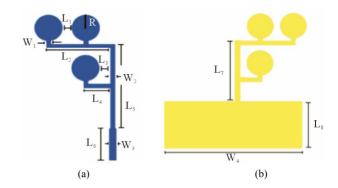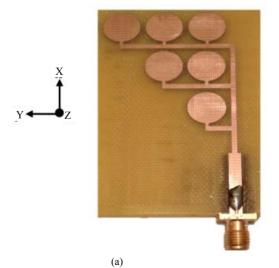
Figure 4. Variables corresponding to each dimension of the proposed dual-band antenna: (a) Top layer; (b) Bottom layer.

Table 2. Dimensions for the proposed dual band antenna at 3.3 - 3.6 and 5.0 - 6.0 GHz.

Variable	Value (mm)
L_1	1.20
L_2	17.8
L_3	1.08
L_4	6.60
L_5	29.3
L_6	10.0
L_7	21.6
L_8	17.8
W_1	1.02
W_2	1.52
W_3	2.54
W_4	49.1
R	5.21
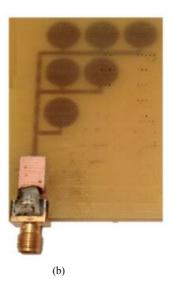
Figure 5. Photo of the fabricated single-band antenna: (a) Top layer; (b) Bottom layer.
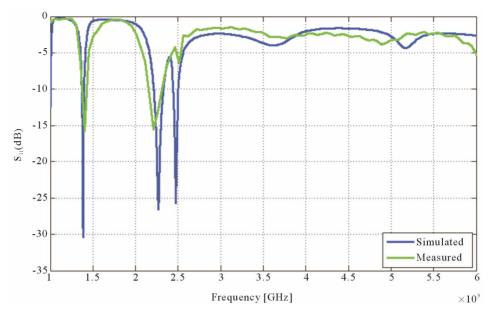
Figure 6. Simulated and measured return loss for the proposed single-band antenna.
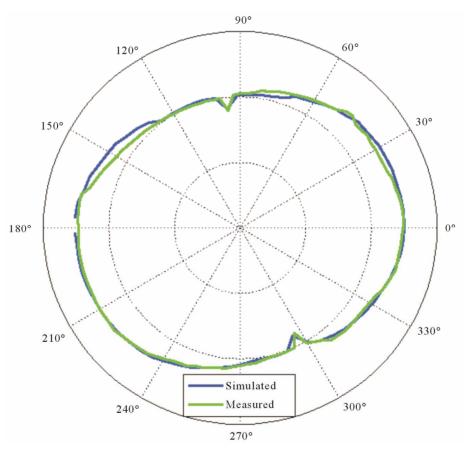
Figure 7. Simulated and measured radiation pattern in xy-plane (coordinate system shown in Figure 5) at 2.45 GHz.

3.2. Dual-Band Antenna at 3.3 - 3.6 and 5.0 - 6.0 GHz

The antennas were fabricated using LPKF Protomat [15]

on double-sided FR-4 materials. The photos of the fabricated dual band antenna are shown in **Figure 8** which has a size of 6.6 × 5.2 (in cm).

Figure 9 shows the comparison between the simulated
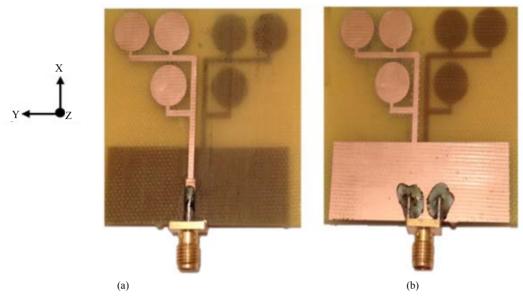
Figure 8. Photo of the fabricated dual-band antenna: (a) Top layer; (b) Bottom layer.
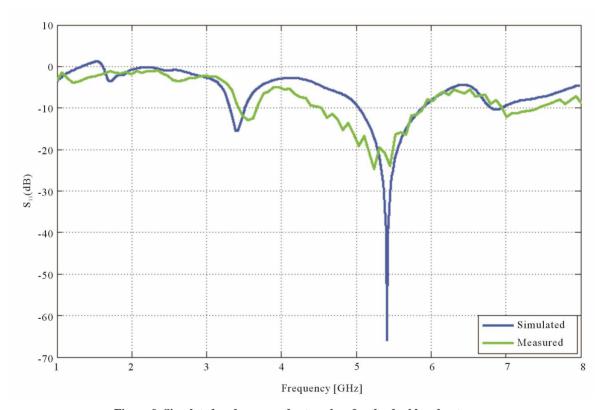
Figure 9. Simulated and measured return loss for the dual band antenna.

and the measured S_{11} results.

The measured S_{11} shows dual band near the designed bands with S_{11} values below −10 dB for both bands. The simulated and measured results give fairly good agreement, and the differences are due to board loss and fabrication errors.

Figure 10 shows the comparison between the simu-

lated and measured radiation pattern in xy-plane at 3.45 and 5.5 GHz which is close to omnidirectional pattern.

4. Conclusion

A microstrip circular antenna arrays were presented for single band at 2.45 GHz and dual bands at 3.3 - 3.6 and

(a)

(b)
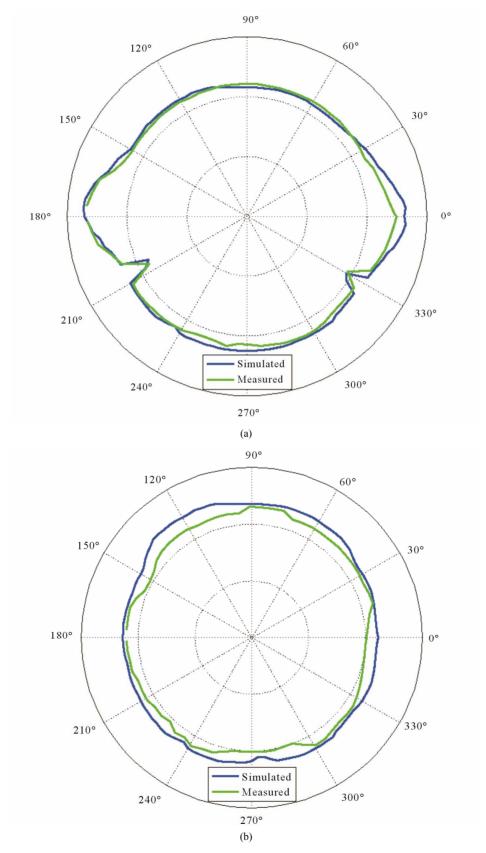
Figure 10. Simulated and measured radiation pattern in xy-plane (coordinate system shown in Figure 8) at (a) 3.45 GHz and (b) 5.5 GHz.

5.0 - 6.0 GHz for WLAN/WiMAX applications. Both antennas were designed with ADS, fabricated on a FR-4 microstrip material, and characterized. Both single band (single sided) and dual band (double-sided) antenna arrays provided omnidirectional pattern with desired gain.

REFERENCES

[1] C.-Y. Pan, T.-S. Horng, W.-S. Chen and C.-H. Huang, "Dual Wideband Printed Monopole Antenna for WLAN/WiMAX Applications," *IEEE Antennas and Wireless Propagation Letters*, Vol. 6, 2007.

[2] A. C. Rao and R. Pandeeswari, "A CPW-Fed Antenna for Dual Band WiMAX/WLAN Applications," *IEEE International Conference on Recent Trends in Information Technology ICRTIT*, 2011.

[3] H.-Y. Lai, Z.-Y. Lei, Y.-J. Xie, G.-L. Ning and K. Yang, "UWB Antenna with Dual Band Rejection for WLAN/WiMAX Bands Using CSRRs," *Progress in Electromagnetics Research Letters*, Vol. 26, 2011, pp. 69-78. doi:10.2528/PIERL11070112

[4] D. Parkash and R. Khanna "Design of a Dual Band Monopole Antenna for WLAN/WiMAX Applications," *IEEE Wireless and Optical Communications Networks*, 2010.

[5] L. Wang, K. Wei, J. Feng, Z. Zhang and Z. Feng, "A Wideband Omnidirectional Planar Microstrip Antenna for WLAN Applications," *IEEE*, 2011.

[6] J. Li, "An Omnidirectional Microstrip Antenna for WiMAX Applications," *IEEE Antennas and Wireless Propagation Letters*, Vol. 10, 2011.

[7] O. Tze-Meng and T. K. Geok, "A Dual-Band Omni-Directional Microstrip Antenna," *Progress in Electromagnetics Research*, Vol. 106, 2010, pp. 363-376. doi:10.2528/PIER10052411

[8] M. B. Bicer and A. Akdagli, "A Novel Microstrip-Fed Monopole Antenna for WLAN/WiMAX Applications," *Journal of Electromagnetic Waves and Applications*, Vol. 26, No. 7, 2012, pp. 904-913. doi:10.1080/09205071.2012.710372

[9] N. AbWahab, Z. Bin Maslan, W. N. W. Muhamad and N. Hamzah, "Microstrip Rectangular 4 × 1 Patch Antenna at 2.5 GHz for WiMAX Application," *2nd International Conference on Computational Intelligence*, *Communication Systems and Networks*, 2010.

[10] Z. Zhong-xiang, C. Chang, W. Xian-liang and F. Minggang, "A 24 GHz Microstrip Array Antenna with Low Side Lobe," Springer, Berlin, 2012.

[11] A. K. Sahu and M. R. Das, "4 × 4 Rectangular Patch Array Antenna for Bore Sight Application of Consial Scan S-Band Traching Radar," *Antenna Week* (*IAW*), 2011.

[12] J. Das, T. A. Khan and M. K. Pal, "Rectangular Patch Antenna Array for Wireless Application," *International Journal of Engineering Science and Technology*, 2012.

[13] T. I. Huque, A. Chhowdhury, K. Hosain and S. Alam, "Performance Analysis of Corporate Feed Rectangular Patc Element and Circular Patch Element 4 × 2 Microstrip Array Antennas," *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 7, 2011, pp. 74-79.

[14] http://www.home.agilent.com/en/pc-1297113/advanced-design-system

[15] http://www.lpkf.com/protomat

Scientific
Research

Back to Back Combined Single Feed Proximity Coupled Antenna with Dumbbell Shaped DGS

Ashwini Kumar Arya, Amalendu Patnaik, Machavaram V. Kartikeyan

Department of Electronics & Computer Engineering, IIT Roorkee, Roorkee, India.
Email: ashwiniarya.iitr@gmail.com, {apatnaik, kartik}@ieee.org

ABSTRACT

Use of defected ground structure (DGS) to reduce the size of patch antenna is presented in this paper. In order to get a dipole like radiation pattern for some specific application a dumbbell shaped DGS is used in the common ground plane of back to back combined single fed proximity coupled antenna. A size reduction of about 60% is achieved. Parametric analysis has been done to see the resonance behavior of the antenna with DGS.

*Keywords***:** *Defected Ground structure, Microstrip Antennas, Proximity Coupling*

1. Introduction

The continuous shrinking size of electronic equipments demands similar size antenna elements in order to fit properly in wireless devices without compromising the other radiation properties of the antenna. In this respect microstrip patch antennas are quite an obvious choice because of its other benefits like low profile, light weight, low cost and easy fabrication.

But as far as size of these patches concerned, the patch length should be around half-a-wavelength for the structure to act as a good radiator. Different techniques have already been used for the antenna size reduction such as using the substrate with high dielectric constant [1] , edge shorted patches with shorting plates or shorting walls, use of the shorting pin at the suitable position etc [2,3].

As far as our understanding goes much has not been reported regarding the use of DGS for size reduction of microstrip antennas, although its application have been reported for harmonic reduction [4], cross-polarization suppression [5] and mutual coupling reduction [6] in antenna arrays etc. Although the back to back geometries have been reported by the various researchers [7,8] but here a new coupling method *i.e.* proximity coupling with the defected ground structure is used for the consideration of the increased bandwidth.

This paper presents the application of DGS for size reduction of microstrip antennas. A dumbbell shaped DGS is used in the common ground plane of a back to back combined single feed proximity coupled microstrip antenna.

2. Defected Ground Structure

Recently there has been an increasing interest in the use of DGSs for performance enhancement of microstrip antennas and arrays. These are realized by etching off a simple shape defect from the ground plane of the planer circuits.

Although various complicated DGSs were reported in the literature, but the simplest one is the dumbbell shaped DGS. **Figure 1(a)** shows the simple and mostly used dumbbell shaped DGS that is etched in the ground plane below the microstrip line, in which both the areas (L_g*W_g) and the slot gap (*g*) play a very important role to find the resonance behavior of the DGS.

The head areas (L_g*W_g) is very useful for the variation in the inductance (L) and the slot (*g*) produces the capacitance (C). The L and C may be calculated from the formulae given below [9].
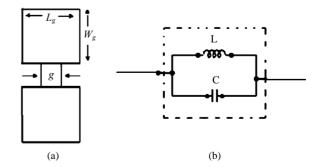
Figure 1. (a) Dumbbell shaped DGS, and (b) DGS Equivalent Circuit.

$$L = \frac{1}{4\pi^2 f_0^2 C} \qquad (1)$$

$$C = \frac{f_c}{2Z_0} \times \frac{1}{2\pi(f_0^2 - f_c^2)} \qquad (2)$$

When this DGS is applied to the antenna, the equivalent inductive part due to the DGS increases and produces equivalently the high effective dielectric constant, thereby decreasing the resonant frequency when the DGS is incorporated in the ground plane of a micro strip antenna.
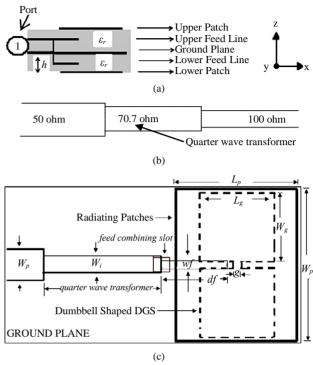
3. Antenna Design with DGS

In this section, the design approach and the performance of the basic antenna and the antenna with DGS is described. At the outset, the single patch antenna was designed and simulated using the CST Microwave studio [10], for the operating frequency at 5.0 GHz. Then another patch of the same size was added in the opposite side of the ground plane and fed in the same way as the first one. The configuration seems as two patch antennas having a common ground plane working at the same frequency. Next the feed lines were combined for the antenna for single feed design. For this purpose the antenna feed lines $W_f = 0.934$ mm with $d_f = 1.5$ mm were designed of 100 ohm and for matching to the 50 ohm transmission line ($W_p = 3.86$ mm) a quarter wave length transformer ($W_t = 2$ mm) was used to give proper matching (**Figure 2**).

The cross-sectional view of the single fed back to back combined proximity coupled antenna is shown in the **Figure 2(a)**. The layouts of matching networks only are emphasized in this **Figure 2(b)** for convenience. It is observed that the antenna designed in this configuration gives the bandwidth of 137 MHz whereas single antenna gives a bandwidth of 67 MHz. This is due to the fact that as the antenna height increases the quality factor decreases and the bandwidth increases. This becomes a multilayer antenna with more height and higher bandwidth as compared to the single patch antenna.

The simple transmission line model was used for the antenna size calculation. The dielectric constant was taken as 3.38 with the loss tangent 0.0025 and of 1.524 mm thickness. The patch lengths L_p and widths W_p are 15 mm and 19 mm respectively. The feed line has been inserted inside the dielectric at a height equal to the half of the height ($h = 3.048$ mm) of the antenna on either side. The dumbbell shaped DGS with dimensions $L_g = W_g = 8.6$ mm and $g = 0.76$ mm was created in the ground plane of the antenna as shown in the **Figure 2(c)**. A small slot was also created for making the antenna with single feed. The two feed lines were connected with a metal strip

which goes through the small slot in the ground plane. The fabricated antenna is shown in **Figure 2(d)**.

(a)

(b)

(c)

(d)
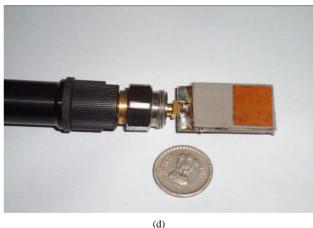
Figure 2. (a) Cross-sectional view of antenna configuration, (b) Feeding Network, (c) Top-View of the antenna, (d) Fabricated Antenna.

4. Results and Discussion

At first the antenna without the DGS in the common ground was simulated and was found to resonate at 5 GHz with 137 MHz Bandwidth.

Then the structure was simulated with the dumbbell shaped DGS. Before reaching to the final size of the DGS, a parametric study was done by varying L_g, W_g and g of

the DGS. As shown in **Figure 3**, with the increase in the value of L_g, the resonant frequency of the antenna is decreasing. Infect, increase in the length of the DGS head gives increasing inductance which in turn decreases the resonant frequency of the antenna.

At this point the increment in the (g) was not possible due to the accuracy in fabrication, so for this reason the other dimension (g) was kept constant for the requirement of the desired frequency (2 GHz). The size of the DGS single square head for the antenna to resonate at 2 GHz (UHF Band) was found to be 8.6 mm × 8.6 mm. The return loss (S11 [dB]) plot of the structure with and without DGS is shown in **Figure 4(a)**. **Figure 4(b)** shows the measured S11 parameter using the HP 8720 B network analyzer. The marker's position near to peak shows the resonance frequency 2.08 GHz with return loss of -13 dB. The result shows good agreement with the simulation results. The measured -10 dB bandwidth is about 60 MHz. The maximum size reduction achieved is about 60%.

(b)

Figure 4. Antenna Return Loss (a) simulated (b) measured.

(— — — measured, ——— simulated) E-Plane
(– – – measured, ▬▬▬ simulated) H-plane
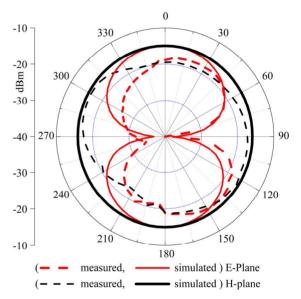
Figure 5. Antenna radiation pattern (measured and simulated) at 2 GHz Frequency.

The measured and simulated power patterns of the antenna are shown in **Figure 5**. It can be observed that the E-plane radiation pattern is similar to the pattern for a dipole antenna. Measurement errors are mainly due to the spurious radiation created by the feeding end and the improper coupling of the elements. However the gain measured experimentally for the proposed antennas with DGS is about -6.9 dB and -7.8 dB (where simulated gain with DGS is 3.8 dB and 5.7 dB is for the antenna without DGS, for both the planes) in both the E and H plane respectively, which is consistent with the size reduction of the antenna.

5. Conclusion

Microstrip patch antenna size reduction with DGS is car-

Figure 3. S11 Vs. Frequency response by varying DGS length (Lg).

(a)

ried out in this work. A dumbbell shaped DGS in the common ground plane of a back to back microstrip structure was found to give a size reduction of about 60% and shifts the resonance frequency from 5 GHz to 2 GHz, with 60 MHz bandwidth facilitating the antenna, to be used for UHF band applications.

REFERENCES

[1]   R. Garg, P. Bhartia, I. Bahl and A. Ittipiboon, "Microstrip Antenna Design Handbook," Artech House, London, 1995.

[2]   P. Salon, M. Keskilammi and M. Kivikoski, "Single Feed Dual Band Planer Inverted-F Antenna with U-Shaped Slot," *IEEE Transactions on Antenna and Propagation*, Vol. 48, No. 8, 2000, pp. 1262-1264. [doi:10.1109/8.884498](doi:10.1109/8.884498)

[3]   D.Nashhat, H. Elsadek and H. Ghali, "Dual-Band Reduced Size PIFA Antenna with U-Slot for Bluetooth and WLAN Applications," *Proceedings of IEEE Antennas Propagation International Symposium*, Columbus, Vol. 2, 22-27 June 2003, pp. 962-965. [doi. 10.1109/APS.2003.1219395](doi.10.1109/APS.2003.1219395)

[4]   H. Liu, X. Sun and J. Mao, "Harmonic Suppression with Photonic Bandgap and Defected Ground Structure for a Microstrip Patch Antenna," *IEEE Microwave and Wireless Component Letters*, Vol. 15, No. 2, February 2005, pp. 55-56. doi:10.1109/LMWC.2004.842809

[5]   D. Guha, M. Biswas and Y. M. M. Antar, "Microstrip Patch Antenna with Defected Ground Structure for Cross Polarization Suppression," *IEEE Antenna and Wireless Propagation Letter*, Vol. 4, No. 1, 2005, pp. 455-458. doi:10.1109/LAWP.2005.860211

[6]   M. Salehi, A. Motevasselian, A. Tavakoli and T. Heidari, "Mutual Coupling Reduction of Microstrip Antenna using Defected Ground Structure," *IEEE Singapore International Conference on Communication System*, Singapore, 2006, pp. 1-5. doi:10.1109/ICCS.2006.301375

[7]   J. Guterman, Y. R. Samii, A. A. Moreira and C. Peixeiro, "Quasi-Omni Directional Dual-band Back-to-Back E-shaped Patch Antenna for Laptop Applications," *Electronics Letters*, Vol. 42, No. 15, 2006, pp. 845-847. [doi:10.1049/el:20061799](doi:10.1049/el:20061799)

[8]   H. Iwasaki, "A Back-to-Back Rectangular-Patch Antenna Fed by a CPW," *IEEE Transactions on Antenna Propagation*, Vol. 46, No. 10, October 1998, pp. 1527-1530. [doi:10.1109/8.725285](doi:10.1109/8.725285)

[9]   D. Ahn, J. S. Park, C. S. Kim, J. Kim, Y. Qian and T. Itoh, "A Design of the Low-Pass Filter Using The Novel Micro strip Defected Ground Structure," *IEEE Transactions on Microwave Theory and Technology*, Vol. 49, No. 1, 2001, pp. 86-91. [doi:10.1109/22.899965](doi:10.1109/22.899965)

[10] EM Simulation Software CST Studio Suite™, v9.

Scientific Research

Vehicular Radio Scanner Using Phased Array Antenna for Dedicated Short Range Communication Service

Tapas Mondal[1], Priyanka Shishodiya[1], Rowdra Ghatak[2], Sekhar Ranjan Bhadra Chaudhuri[3]

[1]Department of Electronics & Communication Engineering, Dr. B. C. Roy Engineering College, Durgapur, India; [2]Microwave and Antenna Research Laboratory, Department of Electronics and Communication Engineering, National Institute of Technology Durgapur, Durgapur, India; [3]Department of Electronics & Telecommunication Engineering, Bengal Engineering and Science University, Shibpur, India.
Email: tapas2k@gmail.com

ABSTRACT

Now a day's accidents are very common due to increased population of vehicle. In order to ensure safety measures in the vehicle this paper has proposed some methodologies regarding careful driving by automatically scanning and analyzing the blind spot area of an intelligent mobile vehicle. A vehicular antenna with minimum perturbation is proposed to be fitted on the vehicle and collect information of the concern area which would ensure visibility of the operator *i.e.* masked or integrated within the car body. This paper has dealt with the design of Tchebyscheff polynomial based prototype planar microstrip phased array antenna and also redesigned the same when implemented in the body of the car being considered as an electromagnetically large element. Both the design has been experimentally verified with the measurement. The simulated and the measured results in both the cases are found to be in good agreement. More than 11 dB gain was observed at perfectly 30° angles from its broad side direction as desired for blind spot detection with minimum amount of electromagnetic interference inside the car.

Keywords: Intelligent Transportation System; Microstrip Phased Array Antenna; Tchebyscheff Polynomial; Electromagnetic Interference

1. Introduction

Intelligent Transport System (ITS) ensures mobility comfort and safety in transportation system. It also absorbs the hazards due to environmental impact. With the progress of the information processing technology, control systems to minimize accidents for the roadways have also been advanced hence one approach to improve the traffic safety is found and that is automatic collection of data by scanning the blind spot area of the vehicle [1] as shown in **Figure 1**. Many methods were proposed to detect the blind spot area but all of them had certain limitations. Devices like dynamic angling side view mirror [2], side view camera model [3], and shadow or edge features detector [4] were used for detecting blind spot area but their performance was affected during bad weather, fog or mist. Also we know that the mechanical systems, response time is more and the system is prone to wear and tear.

A radio frequency method has been proposed in this paper to scan the blind spot zone efficiently. Four rectangular microstrip antennas (RMSA) are arranged in linear configuration with optimal spacing between the patch elements to construct the phased array radar. A

corporate feed network is used to feed the patch element unequally and a progressive phase shifter is designed with 108° delay elements to tilt the main beam in the desired direction and then the total unit is simulated and experimented after placing it on the car body which is electromagnetically a large element. The antenna works in the Dedicated Short-Range Communication Service (DSRCS) frequency band [5]. The design of the microstrip phased array antenna is discussed in Section 2 and after that to place the antenna; the design of the entire car is given in Section 3. Results and discussion are portrayed in Section 4 along with conclusion in Section 5.

2. Design of Microstrip Phased Array Antenna

The microstrip phased array antenna is designed for Dedicated Short Range Communication Service at 5.88 GHz with dielectric constant of 2.32 and substrate thickness of 0.785 mm. Firstly the dimension of the rectangular patch is computed as 16.322 mm by 19.8 mm using the method outlined in [6]. With computed inset feed length of 3.8 mm and for the above dimensions of the rectangular patch the return loss is found to be −8.09 dB
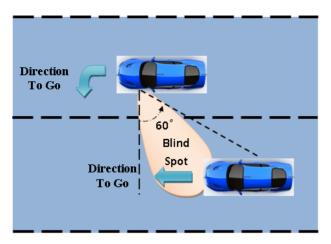
Figure 1. Surrounding regions of a vehicle.

Table 1. Effect on mutual coupling and gain due to variation in spacing (d) between the elements.

d with respect to guided wavelength	d with respect to operating wavelength	Mutual coupling (dB)	Gain (dBi)
$0.6 \lambda_g$	$0.41 \lambda_0$	>10	7.6
$0.7 \lambda_g$	$0.48 \lambda_0$	>13	9.9412
$0.8 \lambda_g$	$0.55 \lambda_0$	>20	10.647
$0.9 \lambda_g$	$0.6 \lambda_0$	>22	10.919

Figure 2. Photograph of the rectangular microstrip antenna element.

with (30.1 + j26.7) ohm impedance at the feed position. The RMSA is simulated and optimized using Ansoft HFSS^{TM}. After optimization the final length is found to be 16.4363 mm with inset feed length of 4.8225 mm keeping the width of the patch unchanged as shown in **Figure 2**.

By considering the above designed patch element, a four element linear array is realized and powered by Tchebyscheff current distribution. The spacing between the elements is kept considering the desired maximum scan angle in order to eliminate the grating lobes within the visible space of the phased array antenna. To optimize the performance of the antenna in respect of its side lobe level (SLL), mutual coupling and gain of the antenna array, the spacing between the elements is studied parametrically [7]. The results of the said study are tabulated in **Table 1**.

After optimization it is found that the optimum spacing between the elements is 0.6 λ_0 while considering, the main lobe to side lobe ratio below 20 dB, optimum mutual coupling and overall gain.

From **Figure 1**, it is observed that the beam of the antenna array is required to be tilted by an angle 30˚ away from the broad side direction. In view of the above a progressive phase shifter of −108˚ is designed with the help of 11.16 mm feed line length. The actual line length is considered as mτ, where m = 0, 1, 2, 3.

To enhance the gain by maintaining the beam width and main lobe to side lobe ratio, the antenna elements are excited by Dolph Chebyshev current distribution. The array consist of four elements, thus third order Tchebyscheff polynomial is calculated. Hence the polynomial is solved and relative current ratio is computed as 1:1.7795:1.7795:1. Both equal and unequal power dividers are designed along with the progressive phase shifter.

For equal power division, a 3 dB equal power divider is designed whose vertical arm is of 50 Ω line and two horizontal quarter-wavelength branch-lines are of 70.71
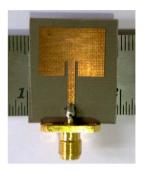
Ω is used [8]. It is shown in **Figure 3(a)**.

To get the unequal current distribution, both the horizontal arm of the microstrip unequal power divider should be of different resistance and it is calculated as 46.368 Ω and 82.49 Ω. The structure is shown in **Figure 3(b)**.

Generally for practical design purpose the gap between the corporate feed networks is taken small but the value of the gap less than 0.11 λ_0 disturbs the frequency response of overall antenna array. The effects of the feed network on the side-lobes are significant in the E-plane pattern. The influence on the H-plane is less important due to the orthogonality of the drive current and the symmetry of all the other currents in this cut [9]. By maintaining the SLL and the surface wave loss, the gaps are taken as 0.11 λ_0 as portrayed in **Figure 4**.

3. Placement of Antenna on the Vehicle

In order to compute the electromagnetic effect of the body of the car on the antenna performance, the entire structure of the car is designed in Ansoft HFSS^{TM}. The complete structure of the car consist of hood, roof, trunk, left and right side doors, left and right side quarter panel for front and rear side which are made up of conducting material but the bumpers and wheels are assigned by layered impedance due to their hard rubber and polyester materials as shown in **Figure 5**.

After going through the typical requirement for scanning the blind spot, the antenna is placed just beside the side view mirror. The following problems in the designed are addressed while simulating the effect of the radiation
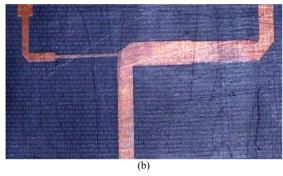
Figure 3. (a) Photograph of the equal microstrip power divider; (b) Photograph of the unequal microstrip power divider.
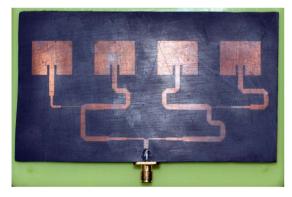
Figure 4. Photograph of the antenna array with corporate feed network.
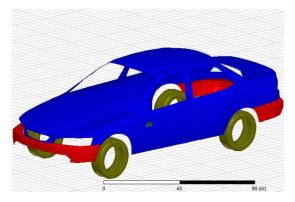
Figure 5. The entire structure of the car along with the antenna array.
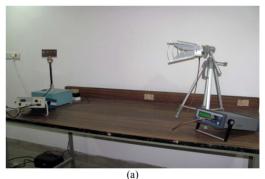
characteristics of the antenna after placing it on the vehicle:

1) Spans a large computational domain compared to the wavelength;

2) Requires a fine grid resolution to resolve the detailed antenna structure;

3) Contains largely non-conformal PEC and dielectric structures.

4. Results and Discussion

For obtaining the scattering parameter characteristics, the microstrip antenna has been simulated over the frequency bandwidth ranging from 5.5 GHz to 6.5 GHz and two different circumstances is observed. The first experiments involve the design, simulation and measurement after fabrication of an individual element and then the array with the same without involving the effect of the car. Both of them should used for optimized performance.

The scattering parameter is measured by Agilent Technology Vector Network Analyzer model no. N5320A (10 MHz - 20 GHz) and the radiation pattern is measured by Hittite HMC-T2100 synthesized signal generator (10 MHz - 20 GHz) and Krytar 9000 B power meter (10 MHz - 40 GHz) with 9530 B power sensor (10 MHz - 20 GHz) as shown in **Figure 6**. From **Figure 7**, it is found
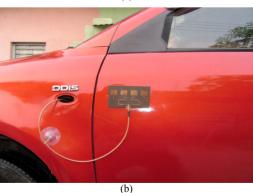
Figure 6. (a) Experimental set-up to measure the radiation pattern of the antenna array; (b) Experimental set-up to measure the radiation pattern of the antenna array after placing it on the car.

(a)

(b)
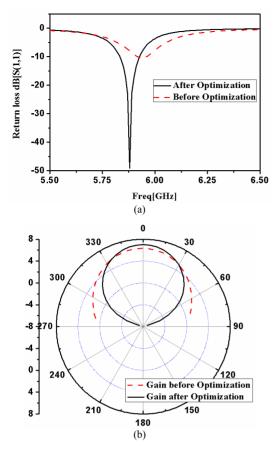
Figure 7. (a) Return loss a single patch before and after optimization; (b) Gain of a single patch before and after optimization.

that after optimization a perfect match is obtained exactly at 5.88 GHz and gain is also improved from 6.33 dB to 7.01 dB with a very good return loss of –47 dB. The current distribution and the radiation pattern of a single RMSA is portrayed in **Figure 8**.

A parametric study of the return loss of antenna array with different spacing between the elements of the corporate feed network is done and is compared in **Figure 9**. The results of the computational model of the entire antenna array with feed network were compared with the measured data and a good agreement is observed in **Figure 10**.

After getting the satisfactory results, next the same array antenna is experimented again with considering the electromagnetic effect of the car. In order to test the antenna array in association with the vehicle, a numerical model of the car is created using Ansoft HFSS^{TM} and the simulated data is compared with the measured in **Figure 11**. **Figure 12** graphically represents the 3D radiation pattern of the antenna array in presence of the computational model of the vehicle. From **Figure 13**, it is clear that the inside electric field strength is not more than 15 Volt per meter which is very less compare to the standard
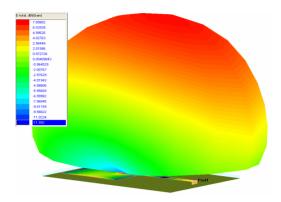
Figure 8. Current distribution and the radiation pattern of the rectangular microstrip antenna.
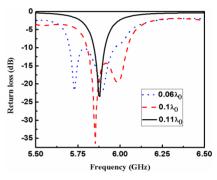
Figure 9. Parametric study of return loss of the antenna array with different spacing between the elements of corporate feed network.

(a)

(b)
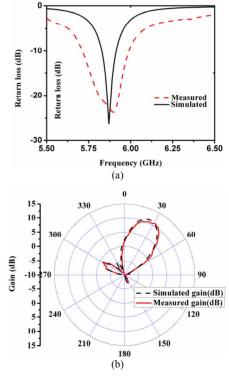
Figure 10. (a) Return loss of the antenna array; (b) Radiation pattern of the antenna array.
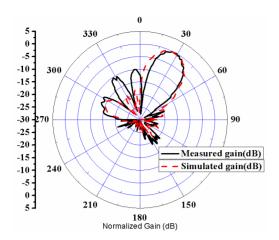
Figure 11. Normalized radiation pattern of the antenna array after placing it on the body of the car.
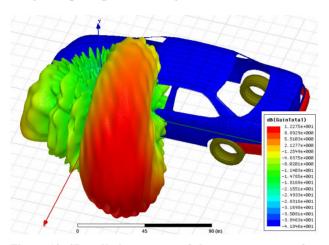
Figure 12. 3D radiation pattern of the antenna array after placing it on a vehicle body.
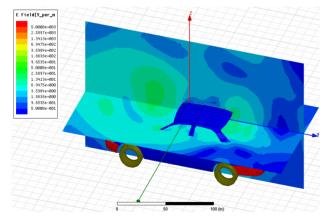
Figure 13. Inside electric field interference of the array after placing it on a vehicle body.

electromagnetic hazards.

5. Conclusion

A Tchebyscheff polynomial based microstrip phased

array antenna is designed to detect the blind spot area of the intelligent mobile vehicle. The design is further simulated in Ansoft HFSS^{TM} after placing the antenna on the body of the car and the results are observed. The result of the computational model of the entire antenna array is compared with the measured data before and after placing the said antenna array on the body of the car and found to be in good agreement. An overall 11 dB gain is obtained while measuring in the desired direction with minimum amount of electromagnetic interference inside the car.

REFERENCES

[1]   M. Griffiths, "Intelligent Transport System," Postnote: Parliamentary Office of Science and Technology, London, 2009. www.parliament.uk/paliamentary_offices/post/

[2]   J. Kuwana and M. Itoh, "Dynamic Angling Side-View Mirror for Supporting Recognition of a Vehicle in the Blind Spot," *International Conference on Control*, *Automation and Systems*, *ICCAS*, Seoul, 14-17 October 2008, pp. 2913-2918.

[3]   S. Jeong, S.-W. Ban and M. H. Lee, "Autonomous Detector Using Saliency Map Model and Modified Mean-Shift Tracking for a Blind Spot Monitor in a Car," *Seventh International Conference on Machine Learning and Applications*, San Diego, 11-13 December 2008, pp. 253-258. doi:10.1109/ICMLA.2008.129

[4]   B.-F. Wu, C.-C. Kao, Y.-F. Li and M.-Y. Tsai, "A Real-Time Embedded Blind Spot Safety Assistance System," *International Journal of Vehicular Technology*, Vol. 2012, 2012, pp. 1-15. doi:10.1155/2012/506235

[5]   A. Tang and A. Yip, "Collision Avoidance Timing Analysis of DSRC-Based Vehicles," *Accident Analysis & Prevention*, Vol. 42, No. 1, 2010, pp. 182-195. doi:10.1016/j.aap.2009.07.019

[6]   T. Mondal, J. S. Roy and S. R. Bhadra Chaudhuri, "Phased Array Antenna Design for Intelligent Transport Systems," *IEEE International Workshop on Antenna Technology*: *Small Antennas and Novel Metamaterial* (i*WAT*), Santa Monica, 2-4 March 2009, pp. 1-4.

[7]   T. Mondal, R. Ghatak and S. R. B. Chaudhuri, "Design and Analysis of a 5.88GHz Microstrip Phased Array Antenna for Intelligent Transport Systems," *IEEE International Sumposium on Antennas and Propagation and CNC*/ *USNC*/*URSI Radio Science Meeting*, Toronto, 11-17 July 2010, pp. 1-4.

[8]   H. J. Visser, "Array and Phased Array Antenna Basics," John Wiley & Sons Ltd., Chichester, 2005. doi:10.1002/0470871199

[9]   E. Levine, G. Malamud, S. Shtrikman and D. Treves, "A Study of Microstrip Array Antennas with the Feed Network," *IEEE Transactions on Antennas and Propagation*, Vol. 31, No. 4, 1989, pp. 426-434. doi:10.1109/8.24162

Scientific
Research

Plus Shape Slotted Fractal Antenna for Wireless Applications

S. Jagadeesha[1], R. M. Vani[2], P. V. Hunagund[3]

[1]S.D.M. Institute of Technology, Ujire Mangalore, India; [2]University Science Instrumentation Center, Gulbarga University, Gulbarga, India; [3]Department of PG Studies and Research in Applied Electronics, Gulbarga University, Gulbarga, India.
Email: Jagadeesh.sd69@gmail.com, vanirm12@rediffmail.com, prbhakar_hunagund@yahoomail.com

ABSTRACT

Fractal antennas are characterized by space filling and self-similarity properties which results in considerable size reduction and multiband operation as compared to conventional microstrip antenna. This paper outlines a multiband antenna design based on fractal concepts. Fractal antennas show multiband behavior due to self-similarity in their structure. The plus shaped fractal antenna has been designed on a substrate of dielectric constant $\epsilon = 4.4$ and thickness 1.6 mm. The proposed antenna is characterized by a compact size and it is microstrip feed fractal patch of order 1/3. It is observed that the antenna is radiating at multiple resonant frequencies. The resonant frequency is reduced from 2.2 GHz to 900 MHz after I & II iterations respectively. Thus considerable size reduction of 81.77% & overall bandwidth of 12.92% are achieved. The proposed antenna is simulated using the method of moment based commercial software (IE3D) and it is found that simulated results are in good agreement with the experimental results.

1. Introduction

In the study of antennas, fractal antenna theory is a relatively new area. The emergence of antennas with fractal geometries has given an answer to two of the main limitations started by Werner (1999) of the classical antennas, which are single band performance and dependence between size and operating frequency. The term "fractal" means broken or irregular fragments. It was originally coined by Mandelbrot (1983) to describe a family of complex shapes that possess an inherent self-similarity or self-affinity in their geometrical structure. Jaggered (1990) defined fractal electrodynamics as an area in which fractal geometry was combined with electromagnetic theory for the purpose of investigating a new class of radiation, propagation and scattering problems. One of the most promising area fractal electrodynamics researches is in its application to antenna theory and design. There are varieties of approaches that have been developed over the years, which can be utilized to archive one or more of these design objectives. The development of fractal geometry came largely from an in depth study of the pattern nature, with the advance of wireless communication system and their increasing importance wide band and low profile antennas are in great demand for both commercial and military applications [1]. A fractal

is a rough or fragmented geometric shape that can be split into parts, each of which is a reduced-size copy of the whole and this property is called self-similarity. Fractal [2] geometries are composite designs that repeat themselves or their statistical characteristics and are thus "self-similar" fractal geometry finds a variety of applications in engineering. Fractal geometry is space filling contours of regular or irregular shapes [3-6], and is super imposed of too much iteration and they describe the self-similar property of fractal geometry [7]. Fractals are a class of shapes which have not characteristic size. Each fractal is composed of multiple iterations of a single elementary shape the iteration can continue infinitely, thus forming a shape within a finite boundary but of infinite length or area. Fractal has the following features 1) It has a finite structure at arbitrarily small scales; 2) It is too irregular to be easily described in traditional Euclidean geometric; 3) It is self-similar; 4) Simple and recursive [8]. Modern telecommunication systems require the antenna with wider bandwidth and smaller dimension than conventionally possible. This has initiated antenna research in various directions, are of which is by using fractal shaped antenna elements. In recent years several fractal geometries have been introduced for antenna application with varying degree of success in improving antenna

characteristics. Some of these geometries have been particularly useful in reducing the size of the antenna, while other designs aim at incorporating multiband characteristics. These are low profile antennas with moderate gain and can be made operative at multiple frequency bands and hence are multifunctional [9]. In our present work we focus on generation of multifrequency which yields increases the bandwidth and size reduction of antenna. A plus shape patch is taken as a base shape and in first iteration four other plus shape patches of the order of 1/3 of base shape are placed touching the base shape. Similarly second iterations are taken by further placing plus shaped patches at even reduced scales. It is found that as the iteration number and iteration factor increases, the resonance frequencies become lower than those of the zero iteration, which represents a conventional plus shape patch.

2. Design Consideration

The base shape of the plus shaped slotted fractal antenna is designed on a dielectric substrate having a relative dielectric constant $\in$ = 4.4 and thickness 1.6 mm as shown in **Figure 1**. This is the reference antenna or base shape antenna. Further this base shape antenna is modified by inserting horizontal slots on both sides with respect to center of patch as shown in **Figure 2** and it is named as antenna1. The length of the slot Ls is varied on either side of the edge as 5 mm, 10 mm, 15 mm, 20 mm, 21.175 mm, 21.675 mm and the frequency variation has been studied. The optimum length obtained is Ls = 21.675 mm *i.e.*, the distance between slots q = 2 mm is considered for further design.

The first iteration patch is designed with four plus shapes of order (1/3) of base shape are placed touching the base shape as shown in **Figure 3** and it is named as antenna 2 and same procedure is repeated for second iteration. This antenna is as shown in **Figure 4** and it is named as antenna 3. For each iteration plus shapes of the order of $(1/3)^n$ of base shape are taken, where n is the number of iterations. The dimension of first iteration can be calculated as

e = (1/3) a & g = (1/3) c also f = (1/3) b & h = (1/3) d.
i = (1/3) e & k = (1/3) g also j = (1/3) f & L = (1/3) h

So with optimized design the dimensions obtained are a = 45.3 mm, b = 15.1 mm, c = 35.4 mm, d = 11.8 mm. The length of the slot is Ls = 21.675 mm and width of the slot Ws *i.e.* r = 2 mm. The dimension of the ground plan is 55 mm × 85 mm. A 50 ohm SMA connector is used to feed the antenna by using microstrip feed technique. Optimized microstripline with following dimension, m = 0.5 mm, n = 18.55 mm, o = 3.05 mm, p = 18.4 mm. The suitable feed location is obtained through optimization process by using the IE3D software. The fabric
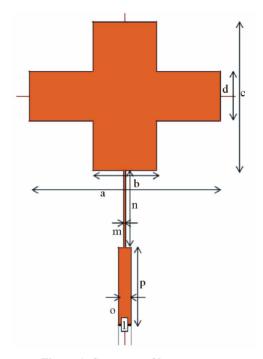
Figure 1. Geometry of base antenna.
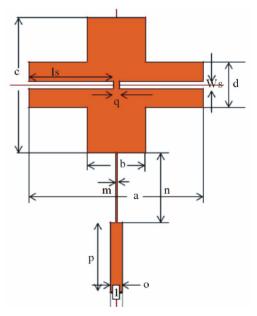
Figure 2. Geometry of antenna 1.

cated photographic view of all proposed antennas is shown from **Figures 5(a)-(e)**.

3. Results and Discussion

The characteristic of the fractal antenna with slot and with iterations has been studied by using IE3D software. Also the results have been verified practically with by using Vector Network Analyzer model Rohde and schewarz, German make ZVK model No.8651.
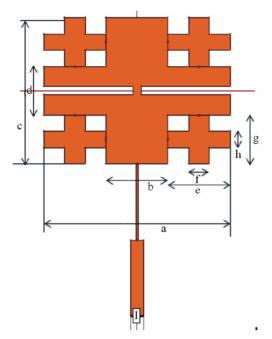
Figure 3. Geometry of base antenna 2.
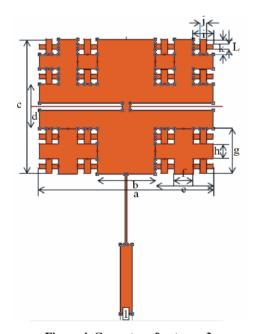
Figure 4. Geometry of antenna 3.

The modified base shape antenna with slots as been optimized by varying slot length Ls. The variation of Ls with resonant frequency of the antenna is shown in **Table 1** and same is presented in graphical form in **Figure 6**. From the tabular results it is found that by increasing slot length Ls on both sides from the edge of the patch, resonant frequency decreases. The lowest possible resonant frequency 1.27 GHz is obtained for Ls = 21.675 mm (*i.e.*, distance between the slots q = 2 mm) & this is taken as optimized length of the slot for further iteration.

The results of all proposed antennas are shown in **Table 2**. The simulated and measured return loss characteristics of proposed antennas are shown from **Figures 7(a)**-**(d)**.

From the results it is clear that the resonant frequency of the antenna 1 *i.e.*, modified base antenna with slot is fr = 1.27 GHz which is lower compared to the base antenna without slot (fr = 2.199 GHz). So the size reduction obtained is 66.85%. The antenna 2 *i.e.* the modified antenna with slots and first iteration gives multiple bands with lower frequency of 0.99 GHz. The size reduction obtained for antenna 2 is 79.88%. Further antenna 3 *i.e.*, with slot and second iteration gives multiple bands with
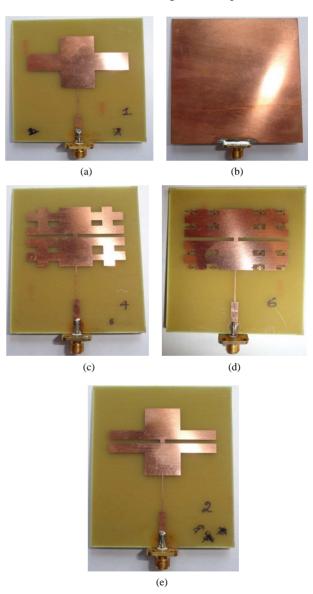
Figure 5. (a) Photograph of top view of base antenna; (b) Photograph of bottom view of base antenna; (c) Photograph of antenna 1; (d) Photograph of antenna 2; (e) Photograph of antenna 3.

Table 1. Variation of slot length v/s resonant frequency.

Slot length (Ls) variation in mm	Resonating frequency (GHz)
0	2.22
5	2.197
10	2.155
15	1.93
20	1.436
12.175	1.34
21.675	1.27
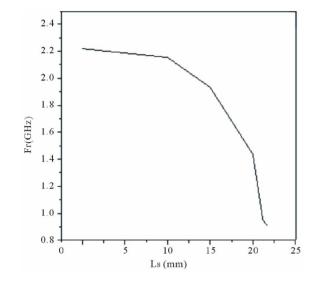
Figure 6. Variation of resonant frequency v/s slot length.

lower resonant frequency of 0.90 GHz. The size reduction obtained for antenna 3 is 81.77% which is more compared to all other proposed antennas.

Further the bandwidths of proposed antennas have been studied through simulation and measurements and the results are shown in **Table 2**. From the results it is clear that the bandwidths of modified antenna with slot, first & second iteration are more compared to base antenna. The measured bandwidth of antenna 1 is 106 MHz (4.895%), antenna 2 is 190 MHz (11.93%) and antenna 3 is 215 MHz (12.92%). The radiation patterns of all proposed antennas are studied and all are giving broadside radiation. Splitting of beam *i.e.* dip is found in **Figure 8(a)** of simulated radiation pattern of base shape at 2.19 GHz. In **Figure 8(b)** splitting of beams merged into broadside pattern at 0.91 GHz at second iterations this results into broadside radiation pattern.

4. Conclusion

This paper presents a new plus shape slotted fractal antenna with first and second iterations. The antenna.3 *i.e.*, slotted fractal antenna with second iteration gives size reduction of 81.77% and band width of 12.92% with broad side radiation pattern. So from the results we conclude that the modified base antenna with slots of second iterations gives a good size reduction and enhanced band width compared to that of modified base antenna with

Table 2. Results of proposed antennas.

Prototype antenna	Resonant frequency fr (GHz)		Return loss (db)		Bandwidth (MHz)		Overall bandwidth (MHz)	
	Sim	Pract	Sim	Pract	Sim	Pract	Sim	Pract
Base antenna	f1 – 2.19	f1 ± 2.19	−19	−24	53	53	79	71
	f2 – 3.42	f2 ± 2.46	−14	−15	26	18		
Antenna1 with q = 2 mm	f1 ± 1.19	f1 ± 1.27	15.8	−12.8	20	18		
	f2 ± 2.52	f2 ± 2.48	18.4	−20.1	113	44	133	106
		f3 ± 2.56		−18.7		44		
Antenna 2 with q = 2 mm (1st itr)	f1 ± 0.96	f1 ± 0.99	16.7	−14.5	22	80		
	f2 ± 2.9	f2 ± 2.97	−14	−26.9	44	50	136	190
	f3 ± 3.08	f3 ± 3.17	21.5	−21	70	60		
Antenna 3 with q = 2 mm (2nd itr)	f1 ± 0.91	f1 ± 0.990	15.2	−14.2	11	87		
	f2 ± 3.07	f2 ± 2.976	19.4	−17.6	70	50	81	215
		f3 ± 3.168		−18.9		78		
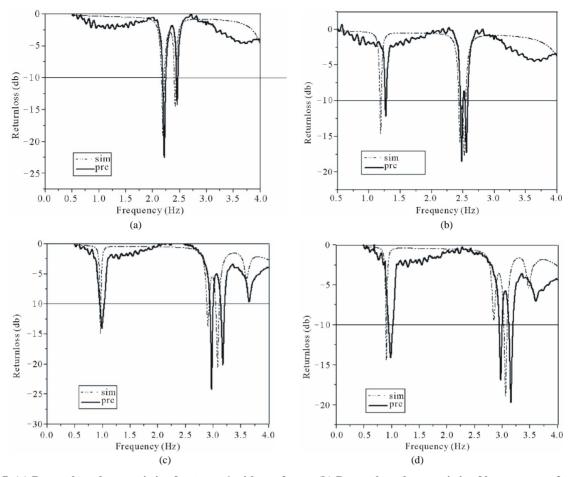
Figure 7. (a) Return loss characteristic of antenna 1 with q = 2 mm; (b) Return loss characteristic of base antenna 2 with q = 2 mm; (c) Return loss characteristic of antenna 2 with q = 2 mm; (d) Return loss characteristic of antenna 3 with q = 2 mm.
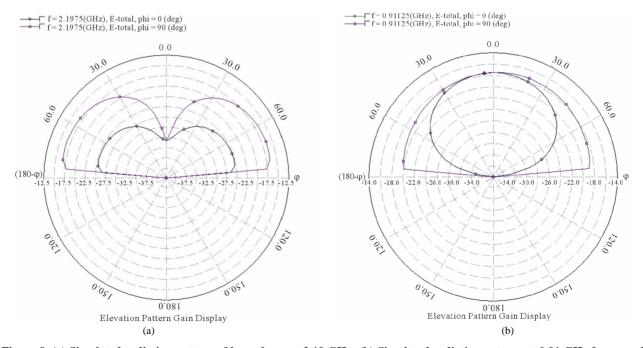
Figure 8. (a) Simulated radiation pattern of base shape at 2.19 GHz; (b) Simulated radiation pattern at 0.91 GHz for second iteration.

slot of first iteration. These antennas may find application in wireless communication systems.

REFERENCES

[1] K. Sing, V. Grewal and R. Saxena, "Fractal Antennas: A Novel Miniaturization Technique for Wireless Communications," *International Journal of Recent Trends in Engineering*, Vol. 2, No. 5, 2009, pp. 172-176.

[2] B. Mandelbrot, "Fractal: Form, Chance and Dimension," W. H. Freeman and Company, San Francisco, 1997.

[3] H. O. Peitgzen, H. Jurgens and D. Sanpe, "Chaos and Fractals," New Frontiers in Science, Springer-Verlag, New York, 1992.

[4] C. Punete, J. Romen, R. Pous, X. Garkia and F. Beitez, "Fractal Multiband Antenna Based on the Serpinski Gasket," *Electronics Letters*, Vol. 32, No. 1, 1996, pp. 1-2. doi:10.1049/el:19960033

[5] J. P. Gianvittorio and Y. Rahmat-Sami, "Fractal Antennas: Novel Antenna: A Novel Antenna Miniaturization Technique and Applications," *IEEE Antenna and Propagation Magazine*, Vol. 44, No. 1, 2002, pp. 20-36. doi:10.1109/74.997888

[6] D. H. Werner and S. Ganguly "An Overview of Fractal Antenna Engineering Research," *IEEE Antenna and Propagation Magazine*, Vol. 45, No. 1, 2003, pp. 38-57.

[7] G. Srivatsun, S. S. Rani and G. S. Krishnan, "A Self-Similar Fractal Cantor Antennas for MICS Based Wireless Applications," *Wireless Engineering Technology*, Vol. 2, No. 2, 2011, pp. 107-111. doi:10.4236/wet.2011.22015

[8] W. Shalan and K. Pahwa, "Multiband Microstrip Rectangular Fractal Antenna for Wireless Applications," *International Journal of Electronics Engineering*, Vol. 3, No. 1, 2011, pp. 103-106.

[9] F. J. Jibrael and M. H. Hammed, "A New Multiband Patch Microstrip Plusses Fractal Antenna for Wireless Applications," *ARPN Journal of Engineering and Applied Sciences*, Vol. 5, No. 8, 2010, pp. 155-158.

New Antenna Array Architectures for Satellite Communications

Miguel A. Salas Natera et al.[*]
Universidad Politécnica de Madrid,
Spain

1. Introduction

Ground stations which integrate the control segment of a satellite mission have as a common feature, the use of large reflector antennas for space communication. Apart from many advantages, large dishes pose a number of impairments regarding their mechanical complexity, low flexibility, and high operation and maintenance costs. hus, reflector antennas are expensive and require the installation of a complex mechanical system to track only one satellite at the same time reducing the efficiency of the segment (Torre et al., 2006).

With the increase of new satellite launches, as well as new satellites and constellation of low earth orbit (LEO), medium earth orbit (MEO), and geostationary earth orbit (GEO), the data download capacity will be saturated for some satellite communication systems and applications. Thus, the feasibility of other antenna technologies must be evaluated to improve the performance of traditional earth stations to serve as the gateway for satellite tracking, telemetry and command (TT&C) operation, payload and payload message or data routing (Tomasic et al., 2002). One alternative is the use of antenna arrays with smaller radiating elements combined with signal processing and beamforming (Godara, 1997).

Main advantages of antenna arrays over large reflectors are the higher flexibility, lower production and maintenance cost, modularity and a more efficient use of the spectrum. Moreover, multi-mission stations can be designed to track different satellites simultaneously by dividing the array in sub-arrays with simultaneous beamforming processes. However, some issues must be considered during the design and implementation of a ground station antenna array: first of all, the architecture (geometry, number of antenna elements) and the beamforming process (optimization criteria, algorithm) must be selected according to the specifications of the system: gain requirements, interference cancellation capabilities, reference signal, complexity, etc. During implementation, deviations will appear as compared to the design due to the manufacturing process: sensor location deviation and sensor gain and phase errors (Martínez & Salas, 2010). In an antenna array, the computation of a close approach of the direction of arrival (DoA) and the correct performance of the beamformer depends on the calibration procedure implemented.

[*] Andrés García-Aguilar, Jonathan Mora-Cuevas, José-Manuel Fernández González, Pablo Padilla de la Torre, Javier García-Gasco Trujillo, Ramón Martínez Rodríguez-Osorio, Manuel Sierra Pérez, Leandro de Haro Ariet and Manuel Sierra Castañer
Universidad Politécnica de Madrid, Spain

This chapter is organized with the following sections. Section 2, introduces the relationship between applications and antenna design architectures. Section 3, introduces the new antenna array architectures for satellite communication including motivation and explains experimental examples. Section 4, explains adaptive antenna array and receiver architectures for adaptive antennas systems considering the beamforming with synchronization algorithms. Finally, Section 5 explains the A3TB concept.

2. Applications and antenna design architectures

In recent effort, new antenna array architectures have been under analysis and development. In (Tomasic et al., 2002) a highly effective, multi-function, low cost spherical phased array antenna design that provides hemispherical coverage is analyzed. This kind of novel architecture design, as the geodesic dome phased array antenna (GDPAA) presented in (Tomasic et al., 2002) preserves all the advantages of spherical phased array antennas while the fabrication is based on well-developed, easily manufacturable, and affordable planar array technology (Liu et al., 2006; Tomasic, 1998). This antenna architecture consists of a number of planar phased sub-arrays arranged in an icosahedral geodesic dome configuration.

In contrast to the about 10 m diameters dome of the GDPAA, there is the geodesic dome array (GEODA) (Sierra et al., 2007) with 5 m diameters dome. This antenna, presented in Fig. 1, has two geometrical structure parts. The first one, is based on a cylinder conformed by 30 triangular planar active arrays, and the second is a half dodecahedron geodesic dome conformed by 30 triangular planar active arrays. The GEODA is specified in a first version for satellite tracking at 1.7 GHz, including multi-mission and multi-beam scenarios (Martínez & Salas, 2010). Subsequently, the system of the GEODA has been upgraded also for transmission (Arias et al., 2010).

a	b	c
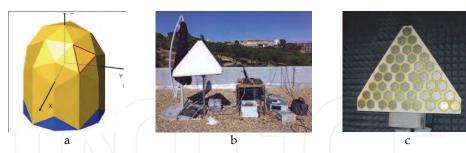
Fig. 1. a) The GEODA, b) The active sub-array demonstration, and c) The 45 elements planar active sub-array.

The antenna arrays technology in the user segment for satellite communications will substitute reflectors providing a more compact and easy to install antenna system, which is an interesting solution e.g. for satellite on the move (SOTM) system. There is a great diversity of solutions for fixed and mobile satellite communication systems including a large number of applications. Inmarsat broadband global area network (Inmarsat-BGAN) (Franchi et al., 2000) is the most representative example among mobile satellite systems (MSS), which gives land, maritime and aeronautical high speed voice and data services with global coverage using GEO satellites at L-band.

MSS services are divided into two groups, those that offer a regional coverage usually with GEO satellites, and those which offer a global coverage based on LEO or MEO satellite constellations. Depending on the coverage, there are some examples for MSS with regional coverage as the mobile satellite system (MSAT) in EEUU, Canada and South America, Optus in Australia, N-Star in Japan, Asia cellular satellite (ACeS) in Asia or Thuraya in the Middle East and in the North of Africa. While for MSS of global coverage there are some examples as Iridium, ICO Global Communications, Globalstar, Teledesic, etc. (Evans, 2009; Wu, 1994). Most of the MSSs work at L and S band, new applications on satellite to mobile terminal links work at X, Ku and Ka band, and satellite to base station connections work at L, S and C band. A number of applications is broad and lead terrestrial telecommunications market to offer a wider coverage: high speed voice and data (internet access, SMS, VoIP), digital video broadcasting by satellite 2 (DVB-S2) and digital video broadcasting satellite services to handhelds (DVB-SH), global position system (GPS) and Galileo, security, control and machinery monitoring on ships and aircrafts, teleeducation or telemedicine.

These modern satellite communications systems require new antenna solutions for base stations, aeronautical applications or personal communications services (PCS) on-the-move (Fujimoto & James, 2001). Within these applications, antenna array systems are potentially the best choice due to, as discussed above, its capability to perform electronically steering or beamforming, increase the antenna gain, and conform over curved or multifaceted surfaces the radiating elements. Portable antennas for PCS must be easy to install and mechanically robust, besides compact and lightweight (García et al., 2010) as the antenna array presented in Fig. 4.a. The design of antenna systems to provide high data rates for reliable PCS boarded on ships is not so strict in term of the geometrical requirements because it does not have space limitations (Geissler et al., 2010). However, in the case of land or airborne vehicles, geometrical and mechanical constraints are more severe. Antennas for terrestrial vehicles must be low profile, and for airborne vehicles aerodynamic shapes must be considered (Baggen et al., 2007; Vaccaro et al., 2010). Moreover, for the civil market conformal antenna arrays (Schippers, 2008; Kanno et al., 1996), or multi-surface arrays (Khalifa & Vaughan, 2007) are suitable choices to deal with the system aesthetic partiality.

Technological challenges have been faced during the implementation of satellite communication systems in the last decades. The design of a Test-Bed flexible and modular for testing or debugging beamforming algorithms and receiver architectures is an invaluable contribution in the educational, research and development area on satellite communication systems. The adaptive antenna array Test-Bed (A3TB) concept is based on the use of antenna arrays with beamforming capability to receive signals from LEO satellites (Salas et al., 2008). The scope of the A3TB is to probe the concept of antenna arrays applied to ground stations instead of reflectors for different applications, such as telemetry data downloading. It is also a good chance for Universities and Research Centers aiming to have their own ground station sited in their installations.

The A3TB ground station relies on the use of an antenna array to smartly combine the received signals from the satellite thanks to the implementation based on software defined radio (SDR) technology. The advantages of the SDR implementation is that A3TB architecture can be used to process any received signal from LEO satellites in the band imposed by the radio frequency (RF) circuits. Moreover, most of the processing is performed in software, so that appropriate routines can be used to process any received signal. The A3TB can be used to analyze the feasibility of different receivers and beamformer

algorithms, regarding the capability to switch the receiver architecture in terms of the synchronizer algorithm configuration (Salas et al., 2007).

The current version has been developed to track The National Oceanic and Atmospheric Administration (NOAA) satellites in the very high frequency (VHF) band, in particular, the automated picture transmission (APT) channel (Salas et al., 2008). Previous versions of A3TB dealt with low rate picture transmission (LRPT) signals from the meteorological operational satellite-A (MetOp-A), where a complete receiver with beamforming and synchronization stages has been implemented (Salas et al., 2007; Martínez et al., 2007).

3. Antenna arrays for satellite communications

Satellite applications require compactness, lightweight and low cost antenna systems to be mounted on a terrestrial vehicle, an aircraft or a ship, or as a portable man-pack or a handset, and to be competitive against ground systems. Its major advantage is the possibility of getting a wider or even a global coverage. For such purposes, antenna arrays offer the technology to get a directive system whose steering direction can be electronically and/or mechanically controlled. However, planar arrays usually cannot steer more than 60°-70° from the normal direction of the antenna (Mailloux, 2005). Thus, when a wider angular coverage is required conformal arrays are an appropriate option (Josefsson & Persson, 2006). Arrays can approximate conformal shapes, such as spheres or cylinders, using several planar arrays, simplifying fabrication of active components (Sierra et al., 2007).

Since the low cost and low weight specifications are of importance, micro-strip antennas are mostly used, due to its capacity to be printed over a dielectric substrate with photolithography techniques. Low cost and low permittivity substrates are usually used such as FR4 or PTFE with different quantities of glass or ceramic impurities. For more demanding applications, ceramics, like alumina or high/low temperature co-fired ceramics (HTCC/LTTC) allow the use of smaller components thanks to its high permittivity, and give robustness against mechanical stresses and high temperatures.

3.1 Geodesic antenna array for satellite tracking in ground station

The aim of using a single antenna for tracking many satellites at the same time avoiding mechanical movements as well as its inexpensive cost make these antennas an alternative to be considered (Salas et al., 2008). Multi-beam ability and interference rejection are facilitated thanks to the electronic control system of such antennas that improves the versatility of the ground stations.

The GEODA is a conformal adaptive antenna array designed for MetOp satellite communications with specifications shown in Table 1. This antenna was conceived to receive signals in single circular polarization (Montesinos et al., 2009). Subsequently, in recent efforts the system has been upgraded also for transmission and double circular polarization (Arias et al., 2010). Hence, operating at 1.7 GHz with double circular polarization it can communicate with several LEO satellites at once in Downlink and Uplink. Current structure is the result of a comprehensive study that valued the ability to cover a given spatial range considering conformal shape surface and a given beamwidth (Montesinos et al., 2009). As Fig. 1 shows, GEODA structure consists of a hemispherical dome placed on a cylinder of 1.5 meters height. Both cylinder and dome are conformed by 30 similar triangular planar arrays (panels). Each panel consists of 15 sub-arrays of 3 elements (cells). The radiating element consists of 2 stacked circular patches with their own

RF circuits. The principal patch is fed in quadrature in 2 points separated 90º in order to obtain circular polarization. The upper coupled patch is used in the aim of improving the bandwidth.

Each panel is able to work itself as an antenna since they have a complete receiver that drives the 1.7 GHz signal to an analog to digital converter (ADC). In order to adapt the signal power to the ADC, it is mandatory to implement a complete intermediate frequency (IF) receiver consisting of heterodyne receiver with an automatic gain control block. Hence, each triangular array has active pointing direction control and leads the signal to a digital receiver through an RF conversion and filtering process. To follow the signal from the satellite, the main beam direction has to be able to sweep an angle of 60º. In this way, it is needed a phase shift in the feeding currents of the single radiating element. Previous calculations have demonstrated that 6 steps of 60 degrees are needed to achieve the required sweeping angle. An adaptive digital system allows the adequate signal combination from several triangular antennas. The control system is explained in (Salas et al., 2010).

Parameter	Specification	Parameter	Specification
Frequency range [GHz] **Tx:** **Rx:**	1.65 to 1.75 1.65 to 1.75	**Isolation between Tx and Rx [dB]**	>20
Polarization	Dual circular for Tx and Rx bands	**VSWR**	1.2:1
G/T [dB/K] **For elevation >30º** **For elevation 5º**	3 6	**SLL [dB]**	-11
EIRP [dBW]	36	**Size [m]**	1.5x1.5x3
3dB beamwidth [deg.]	5	**Accuracy steering [deg.]**	±1.4
Maximum gain [dBi]	29	**Coverage [deg.]:** **Azimuth** **Elevation**	360º >5º
Efficiency [%]	50		

Table 1. Main specifications for GEODA antenna.

3.1.1 Cell radiation pattern

Based on the study presented in (Sierra et al., 2007), the single radiating element is a double stacked circular patch that works at 1.7 GHz with 100 MHz bandwidth. In order to obtain circular polarization, the lower patch, which has 90 mm diameter, is fed by 2 coaxial cables in quadrature. Both coaxial cables connect the patch with a hybrid coupler to transmit and

receive signals with both, right and left, circular polarizations. The upper patch is a circular plate with 78.8 mm diameter, and it is coupled to the lower patch increasing the bandwidth by overlapping both resonant frequencies tuning the substrate thickness and the patch diameter size. Fig. 2.a shows the radiating element scheme and main features of the layer structure are specified in (Montesinos et al., 2009).

A cell sub-array of 3 radiating elements shown in Fig. 2.b is considered the basic module to build the planar triangular arrays. The whole cell fulfills radiation requirements since it has a good polar to crosspolar ratio and a very low axial ratio. Likewise, as it is presented in Fig. 2.c, the radiation pattern shows symmetry and low side lobes for full azimuth.

| a | b | c |
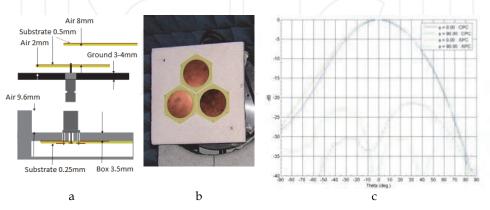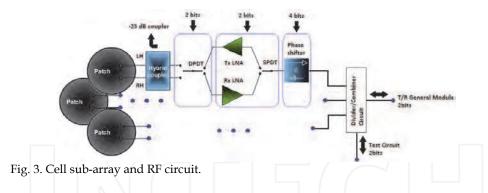
Fig. 2. a) Assembly of the single radiating element, b) Cell scheme, and c) Cell radiation pattern.

3.1.2 Transmission and Reception (T/R) module and cell distribution

Different T/R module configurations have been considered, providing either single or double polarization (Arias et al., 2010). T/R module allows amplifying and controlling the phase shift between signals, received and transmitted, providing an adaptive beam and steering direction controller in the whole working pointing range. As Fig. 3 shows, the design implemented contains a hybrid coupler, enabling double circular polarization; a double pole double throw (DPDT) switch, selecting polarization associated with transmission and reception way; 2 low noise amplifiers (LNAs), which amplify the signal received or transmitted; a single pole double throw (SPDT) switch, choosing transmission or reception way; and phase shifters, introducing multiples of 22.5° relative shift phases to form the desired beam. These surface mount devices have been chosen in order to reduce space and simplify the design.

Signals transmitted/received by the 3 T/R modules placed in a cell are divided/combined thanks to a divider/combiner circuit composed of 3 hybrid couplers that leads the signal to a general T/R module where signal is amplified. Due to transmission and reception duality, 2 SPDT switches are used to select the amplification way. Furthermore, each T/R module has associated a -25dB directional coupler that is used to test T/R modules in the transmission mode. Additionally, reception mode is tested by measuring signal in the divider/combiner circuit. A single pole 6 throw (SP6T) switch selects the path that is tested.

Fig. 3. Cell sub-array and RF circuit.

3.1.3 Control system

The control system has two main parts (Salas et al., 2010), the hardware structure and the control software. The two level hardware structure has the lowest possible number of elements, making the control simpler in contrast to the previous in (Salas et al., 2010). Finally, an inter-integrated circuit (I2C) expander is used to govern T/R modules individually, and one more cover cell needs (LNA of call and test). A multipoint serial standard RS-485 is used to connect the computer with the panels.

3.2 Portable antenna for personal satellite services

New fix and mobile satellite systems (Evans, 2000) require antenna systems which can be portable, low profile and low weight. Planar antennas are perfect candidates to fulfill these specifications. Usually slots (Sierra-Castañer et al., 2005) and printed elements (García et al., 2010) are most used as radiating elements.

3.2.1 Antenna system structure

In this subsection it is introduced a printed antenna for personal satellite communications at X band, in Fig. 4. Table 2 shows main antenna characteristics.

Parameter	Specification	Parameter	Specification
Frequency range[GHz] **Tx:** **Rx:**	7.9 to 8.4 7.25 to 7.75	**Efficiency [%]**	50
Polarization	Dual circular polarization for Tx and Rx bands	**Isolation between Tx and Rx [dB]**	>17
G/T [dB/K]	7	**VSWR**	1.4:1
EIRP [dBW]	32	**SLL [dB]**	-11
3dB beamwidth [deg.]	5	**Size [m]**	40x40x2.5
Maximum gain [dBi]	25	**Weight [Kg]**	2

Table 2. Portable antenna specifications.

This is a planar, compact, modular, low loss and dual circular polarized antenna, for Tx and Rx bands, simultaneously. It is made up by a square planar array of 16x16 double stacked micro-strip patches, fed by two coaxial probes. A hybrid circuit allows the dual circular polarization (Garg et al., 2001). Elements are divided in 16 sub-arrays excited by a global power distribution network of very low losses, minimizing the losses due to the feeding network and maximizing the antenna efficiency. In order to reduce side lobe levels (SLL), the signal distribution decreases from the centre to the antenna edges, keeping symmetry with respect to the main antenna axes. The antenna works at X band from 7.25 up to 8.4 GHz with a 14.7% relative bandwidth for a 1.4:1 VSWR and a maximum gain of 25 dBi.

3.2.2 Sub-array configuration

The sub-array configuration can be seen in Fig. 4.a. It makes possible to separate the fabrication of these sub-arrays from the global distribution network, simplifying the corporative network and getting a modular structure suitable for a serial fabrication process. Each sub-array is a unique multilayer board, where PTFE-Glass substrate of very low losses has been used as base material. The power distribution network is connected to each sub-array through (SMP-type) coaxial connectors.
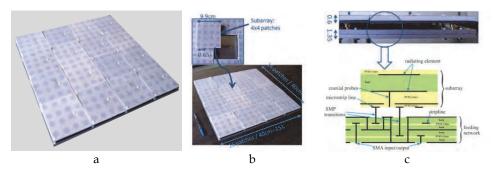
Fig. 4. a) Dual polarized portable printed antenna for satellite communication at X band, b) Sub-array perspective view, and c) Side view and multilayer scheme.

Fig. 5.a and Fig. 5.b show the sub-array unit cell. In order to obtain better polarization purity, each element is rotated 90º and excited by a 90º phase-shifted signal. Moreover, in Fig. 5.c is showed a miniaturized branch-line coupler (BLC) of three branches working as a wide band hybrid circuit (García et al., 2010; Tang & Chen, 2007).
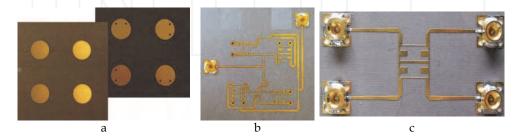
Fig. 5. Unit cell test board, a) Unit cell test board 2x2 stacked patches, b) Micro-strip feeding network, and c) Miniaturized BLC Prototype.

A conventional configuration takes up an area of 13.3 cm² which is big compared to the radiating element and the sub-array subsystem size. Therefore, a miniaturization of the BLC is needed using the equivalence between a λ/4 transmission line and a line with an open-ended shunt stub. An area reduction about 35% is achieved and the hybrid circuit behaves like a conventional BLC. In Fig. 6.b and Fig. 6.c measurement results for the BLC in Fig. 5.c are shown compared with simulations.

Fig. 7 depicts some sub-array measurements. The copular to crosspolar ratio is better than 25 dB and axial ratio is under 0.9 dB in the whole bandwidth.

a   b
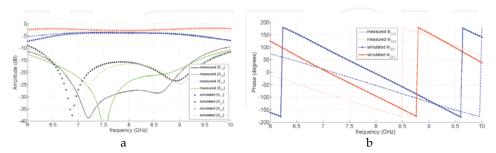
Fig. 6. Miniaturized BLC, Measured and simulated S-parameters in: a) Amplitude, and b) Phase.

a   b
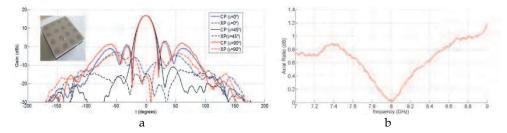
Fig. 7. 4x4 patch sub-array measurements, a) Radiation pattern at 7.75 GHz, and c) Axial ratio for right-handed circular polarization.

3.2.3 Low losses power distribution network

The global feeding network presented in Fig. 8.a is a protected strip-line, where foam sheets of high thickness are used to get low losses. Such a kind of feeding network allows keeping a trade-off between the simplicity of exciting the radiating elements using printed circuits and the loss reduction when the distribution network is separated in a designed structure to have low losses. Losses in the structure are around 0.6 dB/m which yields to 0.3 dB of losses in the line. Two global inputs/outputs using SMA-type connectors, one for each polarization, excite the strip-line networks.

Vertical transitions have to be treated carefully and must be protected to avoid undesired higher order mode excitation. Thereby, it has been design a short-ended pseudo-waveguide, adding some extra losses about 0.3 dB, for two kinds of vertical transitions, as can be seen in Fig. 8.b and Fig. 8.c.
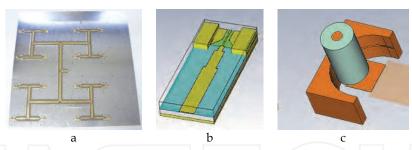
Fig. 8. a) Protected strip-line global corporative network for one polarization, b) Transitions from strip-line to SMA-type connector, and c) Transitions from strip-line to SMP-type connector.

3.2.4 Antenna performance

Fig. 9 depicts measured radiation pattern at 7.75 GHz, gain and axial ratio for the antenna system. It is shown a maximum gain of 25 dBi in the lower band and about 22 dBi in the upper band, and a SLL around 11 dB. Copolar to crosspolar ratio is better than 30 dB and axial ratio is under 0.7 dB. Total losses are about 4 dB in the working band.
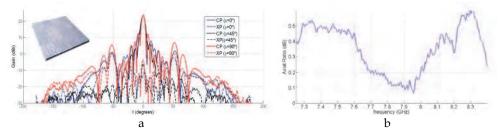
Fig. 9. Antenna measurements results, a) Radiation pattern at 7.75 GHz, and c) Axial ratio for right-handed circular polarization.

3.3 Electronically steerable antennas for mobile and fixed portable systems

At present, two types of electric steerable antenna systems can be used to access the satellite communication services (Bialkwoski et al., 1996). These are: fixed position portable systems and mobile systems such as those installed on a land vehicle. The fixed portable antenna system is relatively easy to be accomplished by the antenna designer. The design involves standard procedures that concern the operational bandwidth, polarization and moderate gain (García et al., 2010). One drawback of the fixed position portable system is that they require the user to be stationary with respect to the ground. This inconvenience can be overcome with the mobile antenna system. A mobile user complicates the scenario since the ground mobile antenna needs to track the satellite (Alonso et al., 1996). The design of such a system is more challenging as new features associated with the mobility of the system have to be incorporated (Fernández et al., 2009). The requirement leads to a narrow beamwidth, for which satellite tracking is required as the vehicle moves around. Electronically steerable antennas enable the development of reconfigurable antennas for satellite applications.

3.3.1 Steerable antenna for fixed position portable systems

This antenna is a fixed satellite communication system with high gain at X band, consisting of an antenna array that integrates 32 2x2 sub-array modules in the complete antenna, as shown in Fig. 10.a. It is a planar and dual circular polarized antenna for Tx and Rx bands simultaneously. It is made up by a planar array of double stacked circular micro-strip patches, fed by 2 coaxial probes to generate circular polarization. A hybrid circuit allows the dual circular polarization as shown in Fig. 10.b.
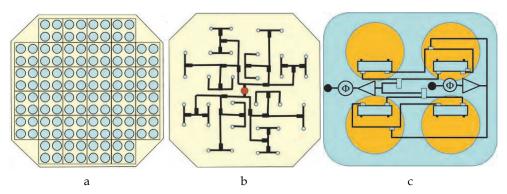
Fig. 10. Active multi-beam antenna, a) Top view, b) Feeding network of the complete antenna, and c) Beamforming network of the 2x2 sub-array module

The antenna has the same design parameters, structure and configuration as the antenna explained in Section 3.2 but with a different feeding network, as previously shown. In this case, the beamforming network requires changes in the feeding phase in the 2x2 sub-arrays, which can be achieved by phase shifters (ϕ) associated with different sub-arrays (Fig. 10.c). All these sub-arrays are connected to a feeding network, in Fig. 10.b, formed by transmission lines with low losses in strip-line. General specifications of the steerable antenna for fixed position portable systems are provided in Table 3.(a).

3.3.2 Automatic steerable antenna for mobile systems

A broadband circularly polarized antenna for satellite communication in X band is presented in Fig. 11 and specified in Table 3.(b). The arrangement features and compactness are required for highly integrated antenna arrays. It is desired to get a low-gain antenna for mobile satellite communications with low speed of transmission. In this system, the antennas are formed by 5 planar 4x4 arrays of antennas, which form a truncated pyramid with a pointing capability in a wide angular range, so that among the 5 planar arrays the complete antenna can cover any of the relative positions between the mobile system and the satellite in a practical way. The scheme of the active antenna can be seen in Fig. 11.

As it can be observed in Fig. 11.a, the antenna terminal is a multi-beam printed antenna shaped as a trunk pyramid capable of directing a main beam in the direction of the satellite. The antenna steering system consists of a multi-beam feeding structure with switches that lets combine the feed of each 4x4 arrays to form multiple beams. Switching the different 4x4 arrays, it is achieved different multiple beams and the variation of the steering direction.

The complete antenna consists of a Tx and Rx module that works independently in the 2 frequency bands.

The antenna has multiple beams covering the entire space to capture the satellite signal without moving the antenna. The signal detected in each of the beams is connected to a switch, which, by comparison, is chosen the most appropriate 4x4 array. The steering direction of the 4x4 array can vary between a range of directions that covers a cone angle range of 90°. To obtain the required gain and cover the indicated range, it is required around 15 beams, which can be obtained by integrating the beamforming networks with switches in the design as presented in (Fernández et al., 2009).

a   b
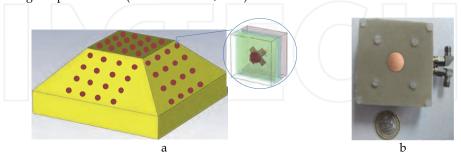
Fig. 11. Complete antenna structure, a) Radiating element of the 4x4 arrays, and b) Prototype top view.

The radiating element of the 4x4 array is one 2 crossed dipoles with a stacked circular patch as shown in Fig. 11.a and Fig. 11.b. In Fig. 12 the cross-section of the radiating element structure is presented.
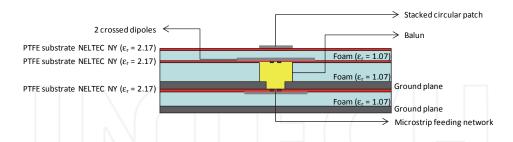
Fig. 12. Cross-section scheme of the radiating element.

The key element of the radiating element feeding structure (Fig. 14.b) is a resonant micro-strip feed ring that has been implemented, as well as a micro-strip 90° branch-line coupler to obtain the desired right hand or left hand circular polarizations (RHCP or LHCP) which ensures adequate port coupling isolation. The S-parameters in amplitude and phase of the micro-strip feeding structure are shown in Fig. 13.a and Fig. 13.b.

Fig. 14.a depicts the S-parameters of the radiating element with the micro-strip feed structure and they fulfill the specification, in Table 3.(b). In Fig. 14.c, the radiation pattern of the radiating element at 7.825 GHz is shown and in Fig. 14.d the radiation pattern of the 4x4

arrays is presented. It is shown a maximum gain of 19.4 dBi at the center frequency band (7.825 GHz). Copolar (CP) to crosspolar (XP) ratio is better than 17 dB and the axial ratio is under -3dB.
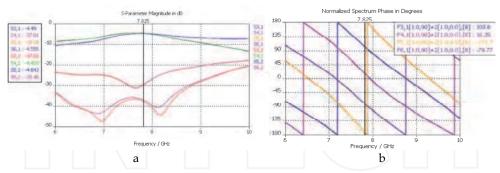
Fig. 13. Micro-strip feeding structure, a) Amplitude of S-parameters, and b) Phase of S-parameters.
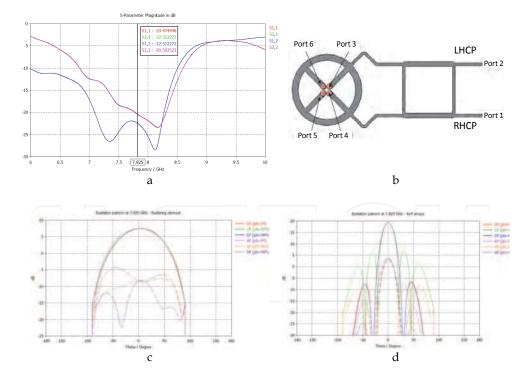
Fig. 14. a) S-parameters, b) Resonant ring + 90° branch-line coupler, c) radiation pattern at 7.825 GHz, and d) 4x4 array radiation pattern.

Parameter	Value (a)	Value (b)	Comments
Freq. range [GHz] Rx Tx	7.25 - 7.75 7.9 - 8.4	7.25 - 7.75 7.9 - 8.4	Microwave applications.
G/T (in Rx) [dB/K]	7	7	
EIRP (in Tx) [dBW]	32	32	
Beamwidth at -3dB [deg.]	4	20	
Polarization	circular	circular	In both, reception and transmission.
Gain [dBi]	>28	>15	
Axial ratio [dB]	< 1	<3	(a) Between ±50°. (b) Between ±45°.
VSWR	< 1.4:1 (-15.6 dB)	< 1.5:1 (-13.9 dB)	
Isolation between ports [dB]	< -17	< -15	
Radiation pattern [deg.]	±35	±90	Steering direction tilt.
Dimensions [cm]	40x40x4	20x20x15	

Table 3. (a) General specifications of the steerable antenna for fixed position portable systems , and (b) General features of the automatic steerable antenna for mobile systems.

3.4 Transmit-array-type lens antenna for terrestrial and on board receivers

Technology in satellite communications has revealed an increasing interest in novel smart antenna designs. Phased-array based designs are basic in electronically reconfigurable devices for satellite applications, which are more and more demanding. The strict requirements in terms of architecture, shape and robustness are important constraints for the development of planar lens-type devices. Regarding the usage and location, lens-type devices are useful for either terrestrial or on board receivers, in vehicular technology. Some clear examples are satellite communications for aircrafts preserving the fuselage aerodynamics or for some other kind of vehicles such as trains, etc.

3.4.1 Introduction to lens-type structures

In a general view, in lens-type a particular signal is received (in our case, an electromagnetic wave with specific features in terms of frequency, wave-front, etc.), it is processed (either complex signal processing techniques or only phase correction tasks can be considered in this interface), and finally, the processed signal is retransmitted.

Regarding the lens configuration, a transmit-array lens consists of three well distinguished interfaces: the first one for signal reception, one interface for signal processing, and the last one for processed signal re-radiation, as depicted in Fig. 15.
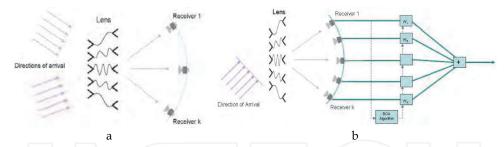
Fig. 15. a) Multi-user scheme with different receivers and transmitters, and b) Adaptive scheme with DoA determination.

These structures are intimately related to reflect-array ones, where the reception and transmission interfaces are turned to be the same interface, with a reflection-type behavior (Encinar & Zornoza, 2001). Although in an equal output phase configuration a transmit-array device behavior would be similar to the one obtained with a reflect-array, the transmit-array offers the advantage of removing the feed blockage.

In a transmission scheme, depending on the transmitter position regarding the lens, a different steering direction is achieved and a different user is pointed. In the case of reception, the situation is the same: the user position configures the direction of arrival, which determines the receiver position around the lens (Padilla et al., 2010a). In adaptive schemes, applying the proper processing algorithm to the signal received in the different receivers around the lens, it is possible to develop an adaptive steering vector, in terms of the desired direction of arrival.

3.4.2 Transmit-array lens architecture and design

Lens-type structures provide two fundamental advantages. First, phase error correction due to spherical wave front coming from the feeding antenna. Fig. 16.a shows this effect. Second, new radiation patterns configuration. Fig. 16.b depicts this fact.
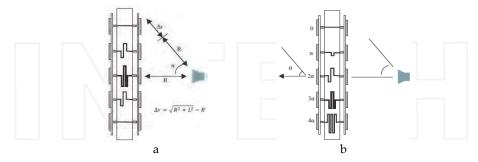
Fig. 16. a) Phase error correction, and b) Radiation pattern reconfiguration.

3.4.3 Electronically reconfigurable devices for active transmit-array lenses

The addition of reconfigurability on transmit-array devices requires the possibility of controlling the phase response of the transmitted signal at each cell of the lens. Electronic control of phase signal may be added in two different ways: First, electronic tuning of the

radiating element phase response (Padilla et al., 2010a): Modifications in the radiating element circuital behavior lead to changes in phase response (arg[S_{21}]). Fig. 17 shows an electronically reconfigurable microwave patch antenna for this purpose, along with the equivalent circuit and prototype outcomes in terms of phase.

Second, electronic tuning of phase shifters in transmission lines (Padilla et al., 2010c): Modifications in the phase response of the phase shifters lead to corresponding changes in phase response. Some options are applied for these devices, such as hybrid couplers, etc. Fig. 18 shows a microwave phase shifter prototype for this purpose, along with the working scheme and its outcomes in phase.
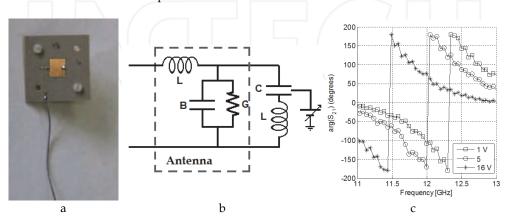
Fig. 17. Electronically reconfigurable antenna, a) Patch antenna prototypes, b) Equivalent circuit, and c) Phase behavior in frequency.
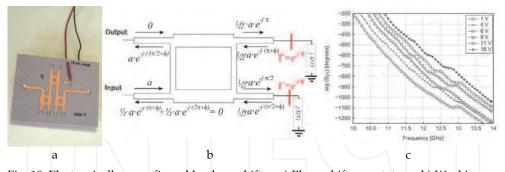
Fig. 18. Electronically reconfigurable phase shifter, a) Phase shifter prototype, b) Working scheme, and c) Phase behavior in frequency.

3.4.4 Electronically reconfigurable active transmit-array prototype

One electronically reconfigurable prototype is presented in Fig. 19 and detailed in this section. The prototype design implies the use of microwave phase shifters according to the design specified in section 3.4.3. This transmit-array lens prototype operates at 12 GHz. Main specifications are provided in Table 4.

Parameter	Value	Comments
Frequency range [GHz]	12 ± 0.5	Microwave applications.
Polarization	Linear	In both, reception and transmission.
Directivity [dBi]	>21	
Axial ratio [dB]	< 1	Between ±50º elevation.
S_{11} [dB]	< -20	
Radiation pattern [deg.]	±30	Steering direction tilt, for both H and V planes.
Feeding antenna [mm]	120	Corrugated horn linearly polarized
Phase shifters [deg.]	360	Full phase range variation.
Transmit-array elements	36	6x6 array topology.
Separation between elements	$0.7\lambda_0$	Related to the wavelength

Table 4. Main features of the electronically reconfigurable transmit-array prototype.

| a | b | c |
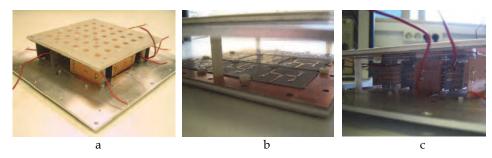
Fig. 19. Transmit-array core, a) Transmit-array prototype, b) Distribution networks, and c) Phase shifter integration.

The electronically controllable steering capabilities are tested and assured for a range of ± 30ºin each main axis. An example of radiation pattern is provided in Fig. 20, for 9º tilt in one of the main axes.

| a | b | c |
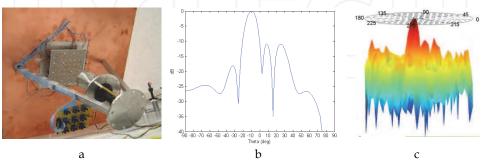
Fig. 20. a) Complete transmit-array with feeder and control circuits; and transmit-array measurement results for 9º tilt in one axis, b) H plane, and c) 3D plot.

4. Adaptive antenna array

Adaptive antennas can be described as systems usually based on three main parts: the antenna array, the receiver architecture and the beamforming scheme. Thus, adaptive antennas have those advantages owing to those three main parts. The system capabilities increase as complexity and development cost do. Furthermore, since signal processing is the basement of the adaptive antenna concept it is important to analyze the design challenges in terms of hardware architecture and components such as processors and embedded systems.

The antenna array provides the capability of performing the antenna pattern meeting the environment requirement under study. Besides, receiver architectures have some interesting advantages depending on the implemented receiver arraying technique such as signal to noise ratio (SNR) and bit error rate (BER) performance enhancement. Furthermore, symbol synchronization and carrier recovery can be used increasing the receiver complexity but providing higher performances. Finally, beamforming schemes use multiple antennas in order to maximize the strength of the signals being sent and received while eliminating, or at least reducing, interference as discussed in Section 4.3.

Adaptive antenna arrays are often called Smart Antennas because they have some key benefits over traditional antennas, by adjusting traffic patterns, space diversity or using multiple access techniques. The main four key benefits are: First, enhanced coverage through range extension by increasing the gain and steering capability of the ground station antenna; Second, enhanced signal quality through multi-target capability and reduction of interferences; finally, adaptive antennas improve the data download capacity in the ground segment of satellite communication by increasing the coverage range (Martínez et al., 2007).

4.1 Design and architecture based on software defined radio

For design there is the well known waterfall life cyclic model (Royce, 1970) that can be used to manage main aspects of the design of architectures. Thus, some tasks must be fulfilled subsequently as follow in Fig. 21.a.

Fig. 21.b shows the design schemes resulting of the requirement analysis stage corresponding software and hardware system specifications. In the depicted scheme, there are some system components such as the radiating element and RF circuits that are often designed under iterative prototyping model.
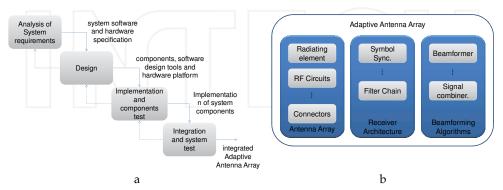
Fig. 21. a) Water life cyclic model of the adaptive antenna array design, and b) Simplified design scheme of adaptive antenna arrays.

Regarding the hardware implementation, tables presented in (Martínez et al., 2007) show the hardware resource consumption in the field programmable gate array (FPGA) Virtex-4 for the least mean squared (LMS) beamforming algorithm with full spectrum combining (FSC) receiver architecture and SIMPLE beamforming algorithm with symbol combining (SC) receiver architecture. Both scheme designs have an antenna array of 2 elements. The algorithm based on correlation requires less hardware. The main difference can be appreciated in the amount of digital signal processing oriented component (DSP48) resources, typically used for filtering applications (Martínez et al., 2007).

4.2 Receiver architectures based on algorithms type

Several receiver architectures can be implemented, and they are frequently based on the type of the beamforming algorithm used. When training signals are available in the transmitted frame, a time-based reference algorithm can be used. However, this solution is only valid when the earth station is capable of demodulating the received training sequence. Other algorithms used in deep space communications are based on signal correlation and they avoid performing the demodulating process. This kind of algorithms are blind techniques that do not require any additional signal demodulation before applying some beamforming technique and work better in low SNR conditions than time-based algorithms. Several receiver architectures can be implemented exploiting the processing capabilities of the SDR, such as FPGA, application-specific integrated circuits (ASICS), and digital signal processing (DSPs). The design of the receiver architecture fundamentally depends on the selection of beamforming algorithms. An example of beamforming technique is the LMS algorithm whose estimation of coefficients or weights requires a temporal reference and is implemented through SC receiver architecture (Fig. 22.a). In the other hand, the SIMPLE algorithm (Rogstad, 1997) constitutes a beamforming technique that is implemented using FSC receiver architecture (Fig. 22.b) in order to perform the calculation of weights.
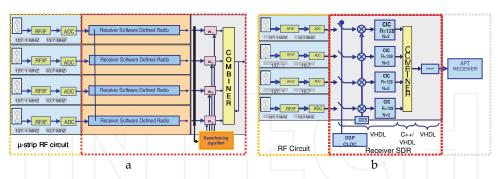
Fig. 22. Comparison of receiver architectures. a) Symbol Combining (SC), and b) Full Spectrum Combining (FSC).

The SC architecture can be divided into two more sub-classes which work on a phase-recovery basis. The complex symbol combining (CSC) recovers the phase information with regard to a reference element using feed-forward and feedback algorithms. One of the advantages of this scheme is that the rate of data sent to the combining module has a rate slightly higher than the symbol rate. For most applications, the symbol rate is relatively low and is a multiple of the data rate. In this kind of schemes, there is an important cost

consideration in real-time applications and the requirements of instrumental phase stability are very severe (Rogstad et al., 2003). Other type of SC architecture is the stream symbol combining (SSC). In this kind of scheme, data are sent to the combining module at a rate equal to the symbol rate. The symbol rate depends on the coding scheme and for most applications is relatively modest. Also, the requirements of instrumental phase stability are no severe, as in the case of CSC scheme. The disadvantage of the SSC is the additional hardware required for each antenna.

Furthermore, there are the baseband combining (BC) and carrier arraying (CA) architectures discussed in (Rogstad et al., 2003). In BC architectures the signal from each antenna is carrier locked and combining in baseband for further demodulation and synchronization. In effect, the carrier signal from the spacecraft is used as a phase reference so that locking to the carrier eliminates the radio-frequency phase differences between antennas imposed by the propagation medium. Besides, in CA architectures, one individual carrier-tracking loop is implemented on each array element. Then, the elements branches are coupled in order to increase the carrier-to-noise ratio (CNR), but losses of radio channel are far compensated (Rogstad et al., 2003).

In general, the selection of the beamforming algorithms is determined by the following aspects: Hardware and computational resources; Speed of convergence and residual error of adaptive algorithms; Calibration requirements and auto-compensation ability; and system signal-transmission characteristics.

4.3 Beamforming techniques for satellite tracking

Some satellites transmit useful information inside its frames for synchronization and tracking purposes. The gathering of satellite data requires the tracking operation along its earth orbit. To accomplish this goal with adaptive array architectures, some beamforming techniques should be implemented. Fig. 23 illustrates a simple example of a narrowband linear adaptive beamformer system.
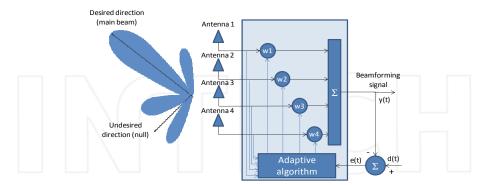
Fig. 23. Adaptive antenna system.

A linear beamformer combines signals according to some weights w_i, to produce a desired radiation pattern. The mathematical expression of a linear beamformer at the array output in vector notation can be expressed as $y = w^H x$, where x is the received signal vector to be combined, w are the weights computed by the beamforming algorithm and H denotes transposition and conjugate of $(\cdot)$.

In adaptive antennas design, weights are dynamically calculated with a certain algorithm in order to optimize some signal parameter like signal to interference-plus-noise ratio (SINR), SNR, or BER. An extended variety of algorithms exist in the literature for beamforming purpose and the most appropriated selection is done depending on the signal characteristics of the received signal.

4.3.1 Blind techniques

Blind beamformers make use of an inherent property of the received signal, such as the ciclo-stationarity of the constant modulus. In the latter, the algorithm eliminates the fluctuation of the signal amplitude and computes the weights to minimize the effect produced by those variations. The algorithms that make use of these methods are denoted as Constant Modulus Algorithms (CMA) (Biedka, 2001).

CMA algorithms present an important disadvantage: as the phase information is not considered, the constellation of quadrature phase shift keying (QPSK) signals commonly used in satellite communications appears rotated after beamforming, which imposes the need of an additional phase recovery subsystem in the array output.

4.3.2 Temporal-reference algorithms

Algorithms based on a temporal reference require a known reference included in the frame of the signal, such as training sequences, unique word (UW) or pilot bits. Thus, these schemes are normally used for digital signals. The aim of these beamformers is the minimization of the energy of an error signal integrated by interferences and noise. In order to reduce the order of the problem, the weight calculation is usually done iteratively.

The most popular adaptive filters are the LMS and Recursive Least Squares (RLS) algorithms (Haykin, 2002). Briefly, the main differences lie in the method to calculate and the final convergence behavior: while LMS has a linear complexity order with the number of antennas in the array, RLS makes use of matrix operation, so that the complexity order is quadratic, but the convergence is faster.

An interesting alternative to the LMS is the Normalized LMS (NLMS), which normalizes the adaptive step to avoid variation during the convergence process. The counterpart is the more intensive processing requirements to calculate signal power and normalization operation.

4.3.3 Correlation-based algorithm

In contrast to beamformers based on temporal reference, schemes based on signal correlation do not require the demodulation of any signal. These techniques are the most popular to extract the spatial information for beamforming, and we have focused on the use of the SIMPLE algorithm (Rogstad, 1997). This algorithm has been used by the Deep Space Network (DSN) of National Aeronautics and Space Administration (NASA) to combine the signals received from spatial probes in radio telescopes located in different sites around the Earth surface. The main disadvantage of correlation based schemes is the lack of ability to cancel interference signals.

4.4 Performance comparison

Some simulation comparisons between spatial and blind algorithms are presented to show benefits and drawbacks. Four algorithms have been selected with a 4-element uniform linear

array (ULA). The spatial algorithms simulated are post-beamformer interference canceller – orthogonal interference beamformer (PIC-OIB) (Godara, 2004) and minimum power distortionless response (MPDR) (Van Trees, 2002). On the other hand, the blind algorithms are the matrix-free EIGEN and the SUMPLE (Rogstad, 1997). The convergence process is compared as a function of the input SNR as depicted in Fig. 24.

As it can be observed from the above results, spatial algorithms outperform blind ones at low SNR, and vice versa. On the other hand, with medium-low SNR and low or absence of interferences, the behavior of all algorithms is quite similar.
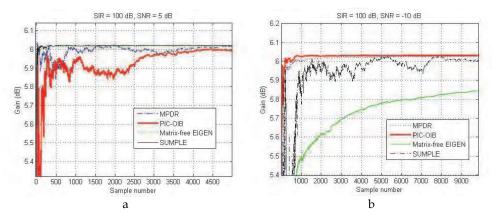
Fig. 24. Convergence behavior of spatial versus blind algorithms in the absence of interferences with several input SNR. a) SNR = 5 dB, and b) SNR = -10 dB.

5. Experimental Test-Bed based on SDR platform

This section presents a test platform known as Adaptive Antenna Array Test-Bed - A3TB, where a comparative study of several beamforming algorithms can be performed and modularity of the architecture is a well proved advantage. The test bed is based on SDR technology and uses a novel architecture that can be used with both blind and spatial-based beamforming algorithms. The A3TB concept can be applied to a number of scenarios as the current version is independent of the signal properties. Simulation results using the A3TB with the APT channel from NOAA satellites show the performance of the concept and the feasibility of the proposed implementation.

The scope of the system development was is to prove the concept of antenna arrays applied to ground stations instead of reflectors for different applications, such as telemetry data downloading or end-user in mobile applications as discussed in the introduction section. In contrast to reflector antennas, antenna arrays offer the possibility of electronic beam-steering avoiding the use of complex mechanical parts and therefore reducing the cost of the antenna. It is also a good chance for Universities and Research Centers aiming to have their own ground station sited in their installations.

5.1 A3TB concept

The A3TB can be defined as *a software-defined radio beamformer applied to a ground station for tracking LEO satellites*. The novelty relies on the use of an antenna array to smartly combine

the received signals from the satellite and its implementation based on SDR technology. The reason to use an antenna array instead of a single antenna is to electronically steer the beam in the direction of the satellite along its orbit without requiring a mechanical system for tracking. In addition to the advantages of the use of SDR technology and antenna array, it is the modularity and flexible architecture implemented in the A3TB. Fig. 25 shows the A3TB architecture where it is evident the feasibility to update or change during operation any of the main blocks. It is possible to change during operation the beamforming algorithm and to include new beamforming modules to the system. Furthermore, changes on the BENADC are possible to implement not during operation, but new receiver architecture at off-line such as those options discussed at follow.

In (Salas et al., 2007), the block diagram represents the software system implementation of the first version of the test-bed prototype and most of it is based on VHDL. Depending on the firmware, three options could be installed into the FPGA Virtex4. The option A is implemented with the signal processing on the PC, so the SIMPLE beamforming is done in the module developed in C++. The option B is implemented completely on VHDL and this option need to export the beamforming weights just to draw the array pattern diagram. Finally, in contrast to the option B, the option C is implemented for the LMS beamforming algorithm.

With the first version of the Test-Bed, the modularity on the selection of firmwares was proved switching between A, B or C receiver architectures, and an important result of the Test-Bed development is the hardware resources occupation presented in (Salas et al., 2007).

The advantage of the SDR implementation is that A3TB architecture can be used to process any received signal from a LEO satellite in the appropriate band imposed by the RF stages. Moreover, most of the processing tasks are performed on software, using appropriate routines to process any receive signal. There are 2 main schemes to implement the beamforming stage: SC and FSC [41]. Both schemes are compared in Section 4.2.

The current version of the A3TB in Fig. 25.a was updated to track NOAA satellites in the VHF band, in particular the APT channel. Previous versions of A3TB dealt with LRPT signals from MetOp-A, where a complete receiver with beamforming and synchronization stages has been implemented(Salas et al., 2007; Martínes et al., 2007).

5.2 Implementation of the A3TB

The A3TB prototype consists of 4 main parts as shown in Fig. 25.a. The first part is the antenna array, which has 4 crossed-dipole antennas as depicted in Fig. 25.b. The second part consists of RF-IF circuits which amplify and down convert to IF incoming signals. Furthermore, an automatic gain control (AGC) was implemented using two steps of variable attenuators in the IF domain.

The third part is the SDR platform which consists of the beamforming algorithms implemented on C++ and the FPGA firmware on VHDL, PC and BENADC blocks show in Fig. 25, respectively. The hardware resources occupation for this Test-Bed implementation is similar to one presented in (Martínes et al., 2007). The last part is the software from weather satellite signal to image decoder (WXtoImg) on the PC using the sound card output/input in order to get the weather satellite image.

Since the implemented architecture is FSC the demodulation is not required and the IF signal is digitized. For the signal processing hardware design the BenADC-v4 has been chosen. This solution includes a FPGA Xilinx Virtex4-SX55 with four 12-bit analog inputs at
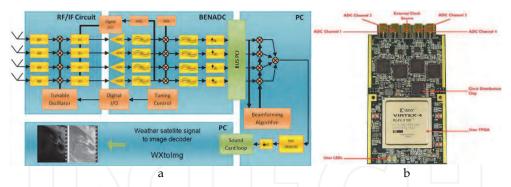
Fig. 25. a) Block diagram of the A3TB, and b) BenADC – Virtex 4-sx55.

250 Msps (Martínes et al., 2007). Digital samples are transferred to the PC where beamforming and subsequent APT demodulation of the array output are performed using C++ routines. This implementation design offers higher flexibility for testing different beamforming schemes. Finally, demodulated APR frames are sent to the WXtoImg software to show meteorological maps.

The A3TB is controlled by the PC for simulations and field trials. The graphical user interface allows presented in (Salas et al., 2008) the user to choose the beamforming algorithm and set all the parameters of the LEO satellite for tracking such as the number of antennas of the array, distance between the elements, direction of arrival and IF frequency. The C++ routine calculates the beamforming weights and plots the synthesized array factor. Subsequently, the reception of meteorological images has real time system requirements. Thus, it is necessary a data transfer from the FPGA to the C++ module to process the samples continuously, and give APT frames to the audio output of the PC. Since, the meteorological satellites often have a low baud rate, in the case of study with NOAA satellites the data transfer is made using two buffers controlled by a thread.

It is important to mention that the A3TB with SDR architecture can evaluate different beamforming algorithms and receiver schemes. The update of A3TB for larger arrays is immediate, as the basis for algorithms is independent of the number of elements in the array. The architecture of a new ground station concept to track LEO satellites based on software defined radio and antenna arraying as Test-Bed is a well proved choice to evaluate future antenna array architectures for satellite communication and benchmark features of the proposed system. As the A3TB VHF version is based on FSC scheme, the concept can be applied to a number of satellite tracing scenarios.

6. Conclusions

The performance analysis of different beamforming algorithms is an important issue in the new generation antenna array development and research. Thus, A3TB helps to analyze beamforming algorithms paving the way for testing and debugging for posteriori use in larger arrays, such as GEODA. Results obtained in real scenarios with A3TB state, for example, that spatial reference algorithms such as MPDR should be used in the absence of interferences, whereas blind algorithms are appropriate for low SNR conditions. Finally, the A3TB can also serve to validate the performance of calibration procedures.

In future work, the A3TB will deal with the system combining of full modularity with the capability of change firmwares based on the first version design of the Test-Bed, plus the flexible architecture of the current design of the Test-Bed based on VHDL, C++ and Antenna Arraying. Furthermore, the addition of more modules to increase the number of antenna array elements is evident in next generations.

7. Acknowledgment

8. References

Torre, A.; Gonzalo, J.; Pulido, M., & Martínez Rodríguez-Osorio, R. (2006). New generation Ground Segment Architecture for LEO satellites. *57th International Astronautical Congress*, pp. 221-226. Valencia, Spain, October 2006.

Tomasic, B.; Turtle, J. & Liu, S. (2002). A Geodesic Sphere Phased Array Antenna for satellite control and communication. *XXVII General Assembly of the International Union of Radio Science,* Maastricht, August 2002.

Godara, L. C. (1997). Application of Antenna Arrays to Mobile Communication, Part II: Beamforming and Direction of Arrival Considerations. *Proc. IEEE ,* vol.85, No.8, (August 1997), pp. 1195-1245.

Martínez, R. & Salas Natera, M. A. (2010). On the use of Ground Antenna Arrays for Satellite Tracking: Architecture, Beamforming, Calibration and Measurements. *61st International Astronautical Congress*, pp. 1-7. Prague, 2010.

Liu, S.; Tomasic, B.; Hwang, S. & Turtle, J. (2006). The Geodesic Dome Phased Array Antenna (GDPAA) for Satellite Operations Support. *IEEE 18th International Conference on Applied Electromagnetics and Communications*, pp. 1-1, Dubrovnik, April 2006.

Tomasic, B. (1998). Analysis and Design Trade-Offs of Candidate Phased Array Architectures for AFSCN Application. *Second AFSCN Phased Array Antenna Workshop*, Hanscom, April 1998.

Sierra Pérez, M.; Torre, A.; Masa Campos, J. L.; Ktorza, D. & Montesinos, I. (2007). GEODA: Adaptive Antenna Array for Metop Satellite Signal Reception. *4th ESA International Workshop on Tracking, Telemetry and Command System for Space Application*, pp. 1-4, Darmstadt, Germany, September 2007.

Arias Campo, M.; Montesinos Ortego, I.; Fernández Jambrina, J. & Sierra Pérez, M. (2010). T/R Module Design for GEODA Antenna. *4th European Conference on Antenna and Propagation,* pp. 115-116, Barcelona, Spain, April 2010.

Franchi, A.; Howell, A. & Sengupta, J. (2000). Broadband mobile via satellite: Inmarsat BGAN. *IEEE Seminar on the Critical Success Factors - Technology, Services and Markets*, pp. 23-30, October 2000.

Evans, J. V. (1998). Satellite systems for personal communications. *Proc. IEEE* , Vol.86, No.7, (July 1998), pp. 1325-1341.

Wu, W. W. (1994). Mobile satellite communications. *Proc. IEEE* , Vol.82, No.9, (September 1994), pp. 1431-1448.

Fujimoto, K. & James, J. (2001). *Mobile Antenna Systems Handbook.* Artech House.

Garcia Aguilar, A.; Inclán Alonso, J. M.; Vigil Herrero, L.; Fernández González, J. M. & Sierra Pérez, M. (2010). Printed antenna for satellite communications. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 529-535, Boston, USA, October 2010.

Geissler, M.; Woetzel, F.; Bottcher, M.; Korthoff, S.; Lauer, A. & Eube, M., (2010). L-band phased array for maritime satcom. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 518-523, Boston, USA, October 2010.

Baggen, R.; Vaccaro, S. & Del Río, D. (2007). Design Considerations for Compact Mobile Ku-Band Satellite Terminals. *EuCAP The Second European Conference on Antennas and Propagation*, pp. 1-5. Edimburg, Scothland, September 2007.

Vaccaro, S.; Tiezzi, F.; Rúa, M. & De Oro, C. (2010). Ku-Band Low-Profile Rx-only and Tx-Rx antennas for Mobile Satellite Communications. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 536-542, Boston, USA, October 2010.

Schippers, H. (2008). Broadband Conformal Phased Array with Optical Beam Forming for Airborne Satellite Communication. *IEEE Aerospace Conference*, pp. 1-17, September 2008.

Kanno, M.; Hashimura, T. T.; Sato, M.; Fukutani, K. & Suzuki, A. (1996). Digital beam forming for conformal active array antenna. *IEEE International Symposium on Phased Array Systems and Technology*, pp. 37-40, October 1996.

Khalifa, I. & Vaughan, R. (2007). Optimal Configuration of Multi-Faceted Phased Arrays for Wide Angle Coverage. *IEEE 65th Vehicular Technology Conference*, pp. 304-308, Boltimore, USA, April 2007.

Salas Natera, M. A.; Martínez Rodríguez-Osorio, R.; Antón Sánchez, A.; García-Rojo, I. & Cuellar, L. (2008). A3TB: Adaptive Antenna Array test-bed for tracking LEO satellites based on software defined radio. *59th International Astronautical Congress*, pp. 313-317, Glasgow. September 2008.

Salas Natera, M. A.; Martínez Rodríguez-Osorio, R. & García-Rojo López, I. (2007). Design of an Adaptive Antenna Array Test-Bed based on Software Radio for Tracking LEO Satellites. *IEEE EuCAP.* Edinburgh, Scotchland, Npvember 2007.

Martínez, R.; Salas Natera, M.; Bravo, A.; García-Rojo, I.; de Haro, L.; Mateo, M. & Gómez, M. (2007). VHF Ground Station with increased angular coverage for reception of meteorological satellites with electronic beamforming. *4th ESA International Workshop on Tracking, Telemetry and Command Systems for Space Application.* Darmstadt, Germany, 2007.

Mailloux, R. (2005). *Phased Array Antenna Handbook,* Artech House, Norwood, Massachusetts, USA.

Josefsson, L. & Persson, P. (2006). *Conformal Array Antenna. Theory and Design*, John Wiley & Sons , Hoboken, New Jersey, USA.

Montesinos, I.; Sierra Pérez, M.; Fernández, J. L.; Martínez, R. & Masa, J. L. (2009). GEODA: Adaptive Antenna of Multiple Planar Arrays for Satellite Communications. *European Conference on Antenna and Propagation.* Berlin, Germany, 2009.

Salas Natera, M. A.; Martínez, R.; De Haro Ariet, L. & Fernández Jambrina, J. (2010). Automated System for Measurement and Characterization of Planar Active Arrays. *IEEE Internationa Symposium on Phase Array Systems and Technology*, pp. 1-6, Boston USA, October 2010.

Sierra-Castañer, M.; Vera-Isasa, M.; Sierra-Pérez, M. & Fernández-Jambrina, J. (2005). Double-Beam Parallel Plate Slot Antenna. *IEEE Transactions on Antennas and Propagation*, vol.53, No.3, (2005) pp. 977-984.

Garg, R.; Bhartia, P.; Bahl, I. & Ittipiboon, A. (2001). *Microstrip Antennas Design Handbook,* Artech House, Norwood, Massachusetts, USA.

Tang, C. & Chen, M. (2007). Synthesizing Microstrip Branch-Line Couplers With Predetermined Compact Size and Bandwidth. *IEEE Transactions on Microwave Theory and Techniques, vol. 55, No.9,* (September 2007), pp. 1926-1934.

Bialkowski, M.; Jellett, S. & Varnes, R. (1996). Electronically Steered Antenna System for the Australian Mobilesat. *IEEE Transaction on Antennas ans Propagation*, vol. 143, No. 4, (August 1996), pp. 347-352.

Alonso, J.; Blas, J.; Garcia, L.; Ramos, J.; Pablos, J. & Grajal, J. (1996). Low Cost Electronically Steered Antenna and Receiver System for Mobile Saltellite Communications. *Trans. IEEE MTT*, vol. 44, No. 12, (December 1996), pp. 2438-2449.

Fernández, J. M.; Rizzo, C. & Sierra-Pérez, M. (2009). Antena Impresa Multihaz con Polarización Circular Derechas/Izquierdas Para Comunicaciones por Satélite en Banda X. *Proccedings of XXIV Simposium Nacional de la Unión Científica Internacional de Radio (URSI).* Santander, Spain, September 2009.

Encinar, J. A. & Zornoza, J. (2003). Broadband design of three- layer printed reflectarrays. *IEEE Transactions on Antennas and Propagation*, vol. 51, No. 7, (July 2003), pp. 1662-1664.

Padilla, P.; Muñoz-Acevedo, A.; Sierra-Castañer, M. & Sierra-Pérez, M. (2010). Electronically reconfigurable transmitarray at Ku band for microwave applications. *IEEE Transactions on Antennas and Propagation*, vol. 58, No. 8, (August 2010), pp. 2571-2579.

Padilla, P.; Muñoz-Acevedo, A. & Sierra-Castañer, M. (2010). Low Loss 360º Ku Band electronically Reconfigurable Phase Shifter. *International Journal of Electronics and Communications,* vol. 64, No. 11, (November 2010), pp. 1100-1104.

Royce, W. W. (1970). Managing the Development of Large Software Systems. *IEEE WESCON*, pp. 328 – 338, 1970.

Rogstad, D. H. (1997). *The SUMPLE Algorithm for Aligning Arrays of Receiving Radio Antennas: Coherence Achivied with Less Hardware and Lower Combining Loss.* TDA Progress Report, Jet Propulsion Laboratory.

Rogstad, D. H.; Mileant, A. & Pham, T. (2003). *Antenna Arraying Techniques in the Deep Space Network.* Deep Space Communication and Navegation Series, JPL California Institute of Technology, Ref. 03-001, Pasadena, USA.

Biedka, T. E. (2001). *Analysis and Development of Blind Adaptive Beamforming Algorithms.* PdD Thesis, Faculty of Virginia Polytechnique Institute and State University, Virginia.

Haykin, S. (2002). *Adaptive Filter Theory* (4th ed.). Prentice-Hall.

Godara, L. C. (2004). *Smart Antennas* (1st ed.). CRC Press..

Van Trees, H. L. (2002). *Optimum Array Processing. Part IV of Detection, Estimation, and Modulation Theory*. Wiley.

Advances in Satellite Communications

Edited by Dr. Masoumeh Karimi

Satellite communication systems are now a major part of most telecommunications networks as well as our everyday lives through mobile personal communication systems and broadcast television. A sound understanding of such systems is therefore important for a wide range of system designers, engineers and users. This book provides a comprehensive review of some applications that have driven this growth. It analyzes various aspects of Satellite Communications from Antenna design, Real Time applications, Quality of Service (QoS), Atmospheric effects, Hybrid Satellite-Terrestrial Networks, Sensor Networks and High Capacity Satellite Links. It is the desire of the authors that the topics selected for the book can give the reader an overview of the current trends in Satellite Systems, and also an in depth analysis of the technical aspects of each one of them.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Miguel Alejandro Salas Natera, Andrés García Aguilar, Jonathan Mora Cueva, José Manuel Fernández, Pablo Padilla De La Torre, Javier García-Gasco Trujillo, Ramón Martínez Rodríguez-Osorio, Manuel Sierra-Perez, Leandro De Haro Ariet and Manuel Sierra Castañer (2011). New Antenna Array Architectures for Satellite Communications, Advances in Satellite Communications, Dr. Masoumeh Karimi (Ed.), ISBN: 978-953-307-562-4, InTech, Available from: http://www.intechopen.com/books/advances-in-satellite-communications/new-antenna-array-architectures-for-satellite-communications

Review Article

Some Recent Developments of Microstrip Antenna

Yong Liu, Li-Ming Si, Meng Wei, Pixian Yan, Pengfei Yang, Hongda Lu, Chao Zheng, Yong Yuan, Jinchao Mou, Xin Lv, and Housjun Sun

Department of Electronic Engineering, School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Yong Liu, fatufo@bit.edu.cn

Although the microstrip antenna has been extensively studied in the past few decades as one of the standard planar antennas, it still has a huge potential for further developments. The paper suggests three areas for further research based on our previous works on microstrip antenna elements and arrays. One is exploring the variety of microstrip antenna topologies to meet the desired requirement such as ultrawide band (UWB), high gain, miniaturization, circular polarization, multipolarized, and so on. Another is to apply microstrip antenna to form composite antenna which is more potent than the individual antenna. The last is growing towards highly integration of antenna/array and feeding network or operating at relatively high frequencies, like sub-millimeter wave or terahertz (THz) wave regime, by using the advanced machining techniques. To support our points of view, some examples of antennas developed in our group are presented and discussed.

1. Introduction

The concept of microstrip antenna was first introduced in the 1950s [1]. However, this idea had to wait nearly 20 years to be realized after the development of the printed circuit board (PCB) technology in the 1970s [2, 3]. Since then, microstrip antennas are considered as the most common types of antennas due to their obvious advantages of light weight, low cost, low profile, planar configuration, easy of conformal, superior portability, suitable for arrays, easy for fabrication, and easy integration with microwave monolithic integrate circuits (MMICs) [4–7]. They have been widely employed for the civilian and military applications such as television, broadcast radio, mobile systems, global positioning system (GPS), radio-frequency identification (RFID), multiple-input multiple-output (MIMO) systems, vehicle collision avoidance system, satellite communications, surveillance systems, direction founding, radar systems, remote sensing, biological imaging, missile guidance, and so on [8].

Despite the many advantages of typical microstrip antennas, they also have three basic disadvantages: narrow bandwidth, low gain, and relatively large size. The narrow bandwidth is one of the main drawbacks of these types of antennas. A straightforward method of improving the bandwidth is increasing the substrate thickness. However, surface wave power increases and radiation power decreases with the increasing substrate thickness [7], which leads to poor radiation efficiency. Thus, various other techniques are presented to provide wide-impedance bandwidths of microstrip antennas, including impedance matching networks using stub [9, 10] and negative capacitor/inductor [11], microstrip slot antennas using the U, L, T, and inverted T slots in the ground plane (sometimes termed defected ground structures (DGSs)) [12, 13], surface wave suppressing using magneto-dielectric substrate [14] and electromagnetic bandgap (EBG) structures [15], and composite-resonator microstrip antennas using metamaterial resonators [16, 17]. Another problem to be solved is the low gain for conventional microstrip antenna element. Cavity backing has been used to eliminate the bidirectional radiation, thereby providing higher gain compared with conventional microstrip antenna [18]. Lens covering is an alternative way to achieve gain enhancement. The lens with canonical profile, like elliptical, hemielliptical, hyper-hemispherical, extended hemispherical, used to focus the radiation beam from the radiator elements. The integrated lens microstrip antenna can be treated as composite antenna combined by microstrip radiator elements and dielectric lens, which is very useful for high frequencies (mm, sub-mm, terahertz (THz), and optical waves) applications

[19]. It is also well known that antenna array is an effective means for improving the gain [20–25].
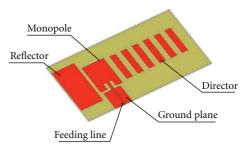
The last limitation of conventional microstrip antennas is the relatively large size, particularly at lower microwave frequencies, since their operation frequencies are related to the electrical size of antenna. In general, the size of the rectangular microstrip antenna should be of order of a half-guided wavelength. This limitation was mathematically studied by Wheeler [26] and Chu [27]. There have been numerous efforts to minimize the antenna size and obtain the electrically small microstrip antenna with the raised demand towards smaller and smaller wireless devices. Inductive or capacitive loading are effective ways to reduce the size of microstrip antennas [28]. In the former work, we demonstrated that the size of microstrip antenna can be miniaturized using composite metamaterial resonators [16, 17]. Magneto-dielectric substrates have been widely used to miniaturize microstrip antennas due to magnetic substrates and could provide wider bandwidths than dielectric substrates [29–32]. Fractal geometries, which are composed by self-similar structures, have opened an alternative way for antenna miniaturization [33].

From the above discussions, we see that many methods and materials are used to improve the properties of microstrip antennas. However, there should be a relationship among bandwidth, gain, and size of the microstrip antennas. Antenna engineers have recognized that the improvement in one antenna property is frequently accompanied by decline in its other performances. For example, the antenna size is reduced usually at the expense of its bandwidth and gain. Therefore, a more comprehensive consideration must be given on further developments of microstrip antennas.

In this paper, we will suggest three areas for further research based on our previous works on microstrip antenna elements and arrays [16–25, 34–41]. We first note that novel microstrip antenna topologies are proposed to meet the desired requirement of variety of potential wireless applications, such as ultrawide band (UWB), high gain, miniaturization, circular polarization, multipolarized, and so on. Next, we discuss the composite antennas based on microstrip antennas which have more potent than each individual antenna. Finally, with the development of micro-/nano-machining techniques, antennas/arrays with highly integration and with highly operating frequencies are discussed. We present some examples of antennas developed in our group to support our points of view.

2. Variety of Microstrip Antenna Topologies

Microstrip antennas have extensively used in commercial and military applications due to their attractive advantages. However, the traditional microstrip antennas have the impedance bandwidth of only a few percent and radiation pattern with omnidirection, which obviously does not meet the requirements of various wireless applications. To this end, a wide variety of microstrip antenna topologies, including different microstrip antenna element structures and different microstrip array arrangements, have been studied to meet the desired requirement such as ultrawide band (UWB), high

(a) The structure of the quasi-Yagi antenna

(b) The photograph of the quasi-Yagi antenna

Figure 1: Compact broad-band quasi-Yagi antenna.

gain, miniaturization, circular polarization, multipolarized, and so forth.

As we know, microstrip antennas inherently have narrower bandwidth and lower gain compared to conventional bulky antennas. Some microstrip antennas with special topologies, like quasi-Yagi, planar reflector antenna, are proposed to replace the conventional bulky antennas. Here, we will take a quai-Yagi antenna as an example to show how to design a planar microstrip antenna with Yagi-Uda end-fire radiation pattern. In addition, a microstrip array with special array topology is designed to get dual-polarized property.

2.1. Compact Broad-Band Quasi-Yagi Antenna. A novel S-band compact quasi-Yagi antenna has been designed, fabricated and measured by our group, as shown in Figure 1. This antenna is composed of a printed monopole-driven element, a printed reflector element, and six printed director elements.

To explain the end-fire radiation behavior of the quasi-Yagi antenna, a comparison of radiation patterns, among (1) microstrip monopole only, (2) microstrip monopole and a reflector, (3) microstrip monopole and a director, (4) microstrip monopole and a reflector with one director, and (5) microstrip monopole and a reflector with six director, is shown in Figure 2. We can observe that both the reflector and the director can increase the end-fire radiation, and it could be substantially improved by increasing the number of directors.

The measured VSWR results are shown in Table 1. A bandwidth of 14% for VSWR less than 1.5 is achieved. The gain of the antenna is above 7.5 dBi, as shown in Table 2. In this design, we see that the microstrip antenna with special topology could be conveniently used to replace the bulky Yagi-Uda antenna.
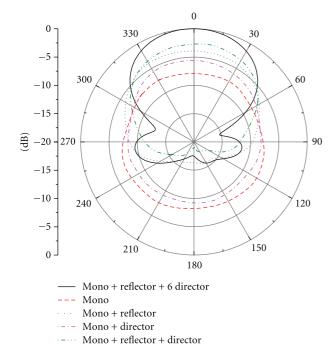
FIGURE 2: Radiation patterns of microstrip monopole only, microstrip monopole and a reflector, microstrip monopole and a director, microstrip monopole and a reflector with one director, and microstrip monopole and a reflector with six director.
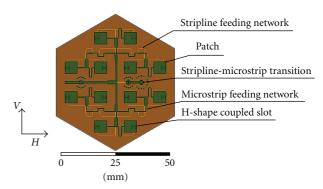
TABLE 1: The measured VSWR of the quasi-Yagi antenna.

No.	Frequency (GHz)			
	3.25	3.5	3.75	Inband
1	1.36	1.34	1.47	<1.5
2	1.37	1.26	1.49	<1.5
3	1.36	1.25	1.48	<1.5
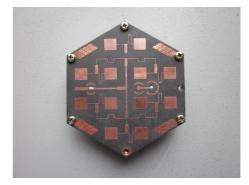
TABLE 2: The measured gain of the quasi-Yagi antenna (unit: dBi).

No.	Frequency (GHz)		
	3.25	3.5	3.75
1	7.57	8.73	8.35
2	7.58	8.55	8.37
3	7.56	8.77	8.51

2.2. Dual-Polarized Microstrip Antenna Array. The dual-polarized antenna is highly required for the radar, electronic countermeasure, and aerospace systems. It is known that the microstrip antenna can easily be integrated with microwave circuits and feeding network. Here, a novel Ku-band dual-polarization microstrip antenna array with a mixed feeding network, that is, the slot coupled feeding (V-port) and the co-plane feeding (H-port), is designed by our group, as shown in Figure 3. It is a three layers structure: top microstrip patch layer, middle stripline feeding network layer, and bottom coplane microstrip feeding network layer. Through proper array arrangement, very good isolation can be obtained.

(a) The structure of the dual-polarized microstrip antenna array

(b) The photograph of the dual-polarized microstrip antenna array

FIGURE 3: Dual-polarized microstrip antenna array.
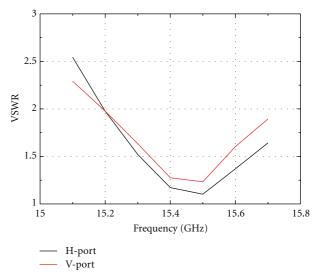
FIGURE 4: The VSWR of the dual-polarized microstrip antenna array.

The VSWR, radiation patterns, and the isolation between two polarizations of the proposed dual-polarized microstrip antenna array are shown in Figures 4, 5, and 6, respectively. The results indicate that this microstrip antenna array has a good impedance matching, good radiation performance, as well as very high isolation (less than −25 dB), which can be an idea candidate for the dual-polarized wireless systems.
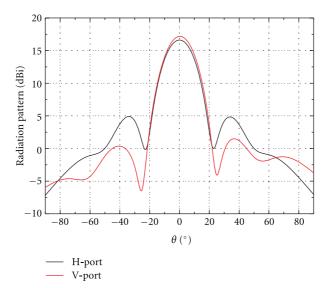
FIGURE 5: The radiation patterns of the dual-polarized microstrip antenna array at the center frequency.
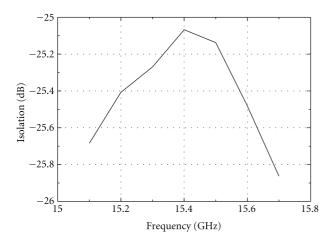
FIGURE 6: The isolation of the dual-polarized microstrip antenna array.

FIGURE 7: The structure of the DCWS.

3. Microstrip-Antenna-Based Composite Antenna

As many antenna designers have found, it is not easy to design an antenna to meet the user-defined stringent performance requirements demanded by special wireless applications like military radars, surveillances, and missile guidance, if only one type of antenna is considered. This difficulty may require the use of two more different types or structures of antenna elements with different characteristics. Composite antenna formed by two more types or structures of antennas is particularly suitable for these applications due to more advantages offered by different types or structures of antennas. For example, it is a challenging task to use single type of antenna to design a dual-band dual-polarization antenna for satellite digital multimedia broadcast (S-DMB) application [36]. A composite antenna composed with a left-handed circularly polarized (LHCP) microstrip antenna and a linear polarized omnidirectional biconical antenna

is proposed by our group to meet this requirement [36]. Another example of composite antenna is comprised of a dielectric lens and microstrip log-period antenna, which has been widely applied to THz systems (this type of antenna will be further discussed in Section 4.2). Here, we will give an example of composite antenna with "structure composite" method.

3.1. Monopulse Circular-Polarized Dielectric Complex Waveguide Slot Antenna Array. Waveguide slot antenna array has been widely used for wireless system, due to its advantages of high radiation efficiency, high power capacity, and high reliability. However, it is hard to overcome the disadvantage of high cost of fabrication.

One composite antenna with waveguide slot antenna array property, termed dielectric complex waveguide slot (DCWS), is composed with slot microstrip line and groove guide, as shown in Figure 7. The slot microstrip line is formed by a metal clad dielectric substrate and slots etched in the metal. This composite antenna not only maintains the advantages of the traditional waveguide slot antenna array but also has the characteristics of high consistence, easy for fabrication, and low cost.

(a) The structure of the monopulse circular-polarized DCWS antenna array (separating view)

(b) The photograph of the monopulse circular-polarized DCWS antenna array.

Figure 8: Ka-band monopulse circular-polarized dielectric complex waveguide slot (DCWS) antenna array.

A Ka-band monopulse circular-polarized dielectric complex waveguide slot (DCWS) antenna array is designed, fabricated, and measured by our group, as shown in Figure 8. It consists of a circular polarization grid, a slot microstrip array, and a groove guide and feeding network. The slot microstrip array is fabricated on a Rogers 5880 film with dielectric constant of 2.2 and the thickness of 0.254 mm. The measured results of VSWR of sum and different port are shown in Figure 9. Figure 10 shows the measured radiation pattern at the center frequency. Some important array performance parameters such as gain, null depth and axial ratio (AR) are also given in Table 3. As shown in the measured results, very good performance can be obtained with the DCWS antenna array. The radiating efficiency of the DCWS antenna array is 80%, which is almost the same as the traditional waveguide slot antenna array. Moreover, the DCWS antenna array has 40% larger bandwidth than the traditional waveguide slot antenna array.

4. Highly Integration and Highly Operating Frequency Antennas Based on Advanced Machining Techniques

It is known that the microstrip antenna was first fabricated using PCB technology in 1970s, nearly 20 years after its
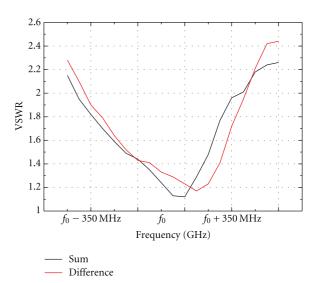
Figure 9: The VSWR of sum and difference port of the monopulse circular-polarized DCWS antenna array.
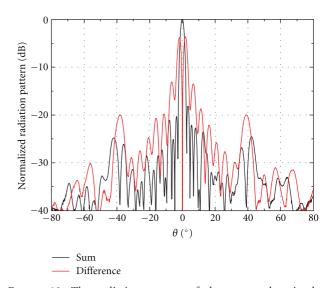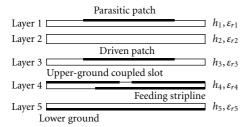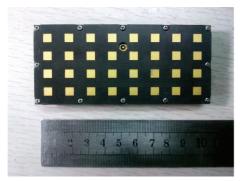
Figure 10: The radiation pattern of the monopulse circular-polarized DCWS antenna array at the center frequency.

concept was first presented in 1950s [1–3]. Clearly, the development of microstrip antennas is closely related with the machining techniques. Recently, various machining techniques, including multilayer printed circuit board (MPCB), complementary metal oxide semiconductor (CMOS), low-temperature cofired ceramics (LTCC), and micro-electro-mechanical systems (MEMS), are highly developed, opening opportunities for innovative antennas, such as active antennas, reconfigurable antennas, metamaterial-based antennas, THz antennas, and so forth. With the availability of high-precision and high-speed advanced machining techniques, microstrip antennas are growing towards highly integration of antenna/array and feed network and operating at relatively high frequencies. Since they are all based on the advanced

(a) Schematic side view of the structure of the high integrate broadband microstrip antenna array

(b) The photograph of the high integrate broadband microstrip antenna array

FIGURE 11: Ku-band high integrate broadband microstrip antenna array using MPCB technology.

FIGURE 12: The VSWR of the high integrate broad-band microstrip antenna array using MPCB technology.

FIGURE 13: The gain of the high integrate broad-band microstrip antenna array using MPCB technology.

TABLE 3: The measured data of the monopulse circular-polarized DCWS antenna array.

Fre. (GHz)	Gain (dBi)	Null depth (dB)	AR (dB)
$f_0 - 0.2$	22.8	-37.3	3.8
f_0	21.9	-29.9	2.9
$f_0 + 0.2$	22.1	-26	4.1

4.1. High Integrate Broad-Band Microstrip Antenna Array Using Multilayer Printed Circuit Board (MPCB) Technology. Recently, with the development of the multilayer printed circuit board (MPCB) technology, the microstrip antennas can be designed and fabricated from one-dimensional (1D) to 2D and even 3D structures.

Based on the MPCB technology, a high integrated broad-band Ku-band microstrip antenna array is designed, fabricated, and measured by our group, as shown in Figure 11. This antenna consists of a parasitic patch, a driven patch, a stripline feeding network, a broad-band coaxial line to stripline transition, some buried screw holes, and some via holes. The feeding network is integrated in the bottom of the substrate of the antenna. As all of the structures fabricated at once, the accuracy and the uniformity can be assured. Two antennas of this type are measured. The measured VSWR, gain, and radiation pattern at the center frequency are shown in Figures 12, 13, and 14, respectively. The measured results show that this antenna maintains good radiation and matching performances with relative bandwidth of 13%. They have also shown good uniformity by using MPCB technology.

4.2. THz Wave Planar Integrated Active Microstrip Antenna Using Micro-Electromechanical Systems (MEMs) Technology. THz waves typically include frequencies between 0.1 THz and 10 THz. THz technology is now becoming a promising technology which has potential applications in many fields, such as short-range communication, biosensor, imaging,

machining techniques, we suggest that a third research area of microstrip antennas is constantly introducing novel advanced machining techniques. In the following, two examples will be presented to show how important the advanced machining technique is to fabricate microstrip antennas. One is the highly integrate broad-band microstrip antenna array fabricated using MPCB technology. Another is THz wave planar integrated active microstrip antenna using MEMS technology.

(a) E-plane

(b) H-plane

FIGURE 14: The radiation pattern of the high integrate broad-band microstrip antenna array using MPCB technology at the center frequency.

(a) The photograph of the THz monolithic antenna

(b) The photograph of the THz monolithic antenna covered by a dielectric lens

FIGURE 15: THz wave planar integrated active microstrip antenna using micro-electromechanical systems (MEMSs).

national security, space exploration and communication, and so forth [39–46]. To realize THz transceiver system, antenna is an essential component. We often use horn antenna, lens antenna, and dielectric parabolic antenna, for THz systems. However, they are not easy to integrate with monolithic integrate circuits. Although the microstrip antenna has the merits of small volume, light weight, and easy integration with circuit, it is difficult to be processed in such high-frequency regions. MEMS technology opens the way to design of THz antennas, circuits, and systems. THz monolithic antenna fabricated using MEMS technology and covered by a dielectric lens, which can be considered a composite antenna, are designed, fabricated, and measured by our group, as shown in Figure 15.

Diodes have the functions of mixing and/or modulating the carrier-wave signal. It is an effective way to reduce the propagation path for detectors application by integrating the diode and microstrip antenna. The extended hyper-hemispherical dielectric lens is used to increase the gain of the microstrip antenna. An antenna-coupled detector integrated with a dielectric lens is designed and fabricated up to THz range by our group. The planar microstrip log-spiral antenna and log-period antenna have been fabricated using micro-electromechanical systems (MEMSs) technology. The photographs of the antennas are demonstrated in Figure 15. The measured responses of the antenna-coupled detector working at different frequency bands are shown in Figure 16, which can be considered to determine the effective operating frequencies [19, 40]. This detector gave a valid response from 12 GHz to 110 GHz frequencies. The results prove the validity and feasibility of the THz antenna designed using micro-electromechanical systems (MEMS) technology.

5. Conclusion

The advantages and disadvantages of microstrip antennas are discussed in this paper. In particular, three areas for further development of microstrip antennas are presented based on our previous works on microstrip antenna elements and arrays. Variety of microstrip antenna topologies and microstrip-antenna-based composite antenna are discussed, and

(a) Ku-band

(b) Ka-band

(c) V-band

(d) W-band
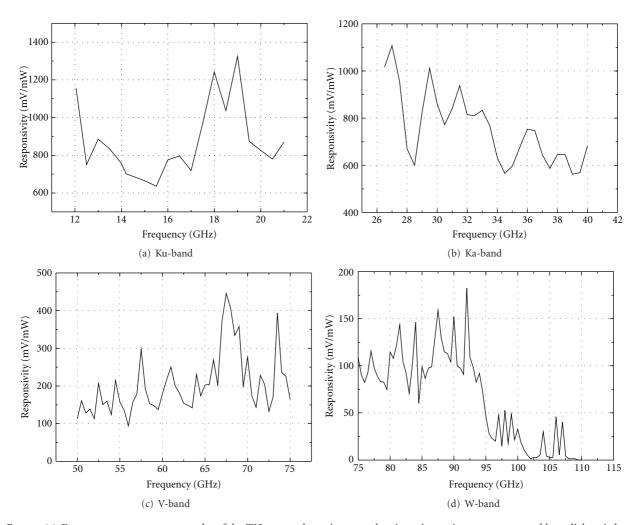
FIGURE 16: Frequency responses test results of the THz wave planar integrated active microstrip antenna covered by a dielectric lens.

the advanced machining techniques pushing the microstrip antennas towards the highly integration of antenna/array and feeding network and the highly operating frequencies are described. To demonstrate the distinctive features of novel microstrip antennas, various antenna elements and arrays for different applications are presented. This paper has shown that the microstrip antennas are still very promising paradigm for civilian and military wireless applications.

References

[1] G. A. Deschamps, "Microstrip microwave antennas," in *Proceedings of the 3rd USAF Symposium on Antennas*, 1953.

[2] E. V. Byron, "A new flush-mounted antenna element for phased array application," in *Proceedings of the Phased Array Antenna Symposium*, pp. 187–192, 1970.

[3] J. Q. Howel, "Microstrip antennas," in *Proceedings of the Digest of the International Symposium of the Antennas and propagation Society*, pp. 177–180, 1972.

[4] I. J. Bahl and P. Bhartia, *Microstrip Antennas*, Artech House, 1980.

[5] J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna Theory and Design*, Peter Peregrinus, 1981.

[6] K. R. Carver and J. W. Mink, "Microstrip antenna technology," *IEEE Transactions on Antennas and Propagation*, vol. 1, no. 1, pp. 2–24, 1981.

[7] D. M. Pozar, "Microstrip antennas," *Proceedings of the IEEE*, vol. 80, no. 1, pp. 79–91, 1992.

[8] R. Garg, I. Bhartia, I. Bahl, and A. Ittipiboon, *Micro-Strip Antenna Design Handbook*, Artech House, Boston, Mass, USA, 2001.

[9] D. M. Pozar and B. Kaufman, "Increasing the bandwidth of a microstrip antenna by proximity coupling," *Electronics Letters*, vol. 23, no. 8, pp. 368–369, 1987.

[10] H. F. Pues and A. R. Van de Capelle, "Impedance-matching technique for increasing the bandwidth of microstrip antennas," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 11, pp. 1345–1354, 1989.

[11] A. Kaya and E. Y. Yüksel, "Investigation of a compensated rectangular microstrip antenna with negative capacitor and negative inductor for bandwidth enhancement," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 5, pp. 1275–1282, 2007.

[12] K. L. Wong and N. H. Hsu, "A broad-band rectangular patch antenna with a pair of wide slits," *IEEE Transactions on Antennas and Propagation*, vol. 49, no. 9, pp. 1345–1347, 2001.
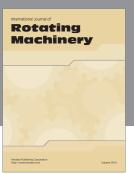
[13] S. I. Latif, L. Shafai, and S. K. Sharma, "Bandwidth enhancement and size reduction of microstrip slot antennas," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 3, pp. 994–1003, 2005.

[14] P. M. T. Ikonen, S. I. Maslovski, C. R. Simovski, and S. A. Tretyakov, "On artificial magnetodielectric loading for improving the impedance bandwidth properties of microstrip antennas," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 6, pp. 1654–1662, 2006.

[15] R. Cocooli, F. R. Yang, T. P. Ma, and T. Itoh, "Apertu-recoupled patch antenna on UCPBG substrate," *IEEE Transactions on Microwave Theory and Techniaues*, vol. 1, no. 47, pp. 2123–2130, 1999.

[16] L.-M. Si and X. Lv, "CPW-FED multi-band omni-directional planar microstrip antenna using composite metamaterial resonators for wireless communications," *Progress in Electromagnetics Research*, vol. 83, pp. 133–146, 2008.

[17] L. M. Si, H. Sun, Y. Yuan, and X. Lv, "CPW-fed com-pact planar UWB antenna with circular disc and spiral split ring resonators," in *Progress in Electromagnetics Research Symposium*, pp. 502–505, 2009.

[18] Y. Yuan, L.-M. Si, Y. Liu, and X. Lv, "Integrated log-periodic antenna for Terahertz applications," in *International Conference on Microwave Technology and Computational Electromagnetics*, pp. 276–279, 2009.

[19] L. Xin, M. Jinchao, Y. Yong, Y. Weihua, N. Hongbin, and G. Yafen, "Integrated active antennas in terahertz focal plane imaging system," in *Proceedings of the International Workshop on Antenna Technology*, pp. 157–163, 2011.

[20] Q. Xu, W. Zhang, H. Sun, and X. Lu, "Millimetre wave multi-polarised microstrip antenna array and application example," *IET Microwaves, Antennas and Propagation*, vol. 4, no. 10, pp. 1525–1530, 2010.

[21] M. Wei, H. Deng, H. Sun, and Y. Liu, "Design of an X/Ka dual-band co-aperture broadband microstrip antenna array," in *Proceedings IEEE International Conference on Microwave Technology and Computational Electromagnetics*, pp. 217–220, 2011.

[22] Y. Liu, X. Lu, Y. Yuan, Y. Chen, and L. Shi, "Integrated design and research of Ka-band electronically large mono-pulse planar antenna and feed system," in *Proceedings of the 7th International Symposium on Antennas, Propagation and EM Theory (ISAPE '06)*, pp. 150–153, October 2006.

[23] Y. Liu, H. J. Sun, X. Lu et al., "Study of the millimeter microstrip monopulse antenna array in the dual-sensor system," in *Proceedings of the 7th International Symposium on Antennas, Propagation and EM Theory (ISAPE '06)*, pp. 157–160, October 2006.

[24] Y. An, X. Lv, and B. Gao, "Developing a kind of mi-crostrip array antenna with beam squint," in *Proceedings of the 5th International Symposium on Antennas, Propagation and EM Theory*, pp. 443–446, 2000.

[25] L. Yong, L. Xin, H. Zhao, and Y. Yuan, "Integrated design and research of novel KA-band circular polarized monopulse interrogator array antenna," in *Proceedings of the IET International Radar Conference 2009*, pp. 1–4, April 2009.

[26] H. A. Wheeler, "Fundamental limitations of small an-tennas," *Proceedings of the IRE*, vol. 35, pp. 1479–1484, 1947.

[27] L. J. Chu, "Physical limitations of omni-directional antennas," *Journal of Applied Physics*, vol. 19, no. 12, pp. 1163–1175, 1948.

[28] R. Azadegan and K. Sarabandi, "A novel approach for mini-turization of slot antennas," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 3, pp. 421–429, 2003.

[29] H. Mosallaei and K. Sarabandi, "Magneto-dielectrics in elec-tromagnetics: concept and applications," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 6, pp. 1558–1567, 2004.

[30] P. M. T. Ikonen, S. I. Maslovski, C. R. Simovski, and S. A. Tretyakov, "On artificial magnetodielectric loading for improving the impedance bandwidth properties of microstrip antennas," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 6, pp. 1654–1662, 2006.

[31] P. M. T. Ikonen, K. N. Rozanov, A. V. Osipov, P. Alitalo, and S. A. Tretyakov, "Magnetodielectric substrates in antenna miniaturization: potential and limitations," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3391–3399, 2006.

[32] P. Ikonen and S. Tretyakov, "On the advantages of magnetic materials in microstrip antenna miniaturization," *Microwave and Optical Technology Letters*, vol. 50, no. 12, pp. 3131–3134, 2008.

[33] J. Anguera, C. Puente, C. Borja, and J. Soler, "Frac-tal-shaped antennas: a review," in *Wiley: Encyclopedia of RF and Microwave Engineering*, K. Chang, Ed., vol. 2, pp. 1620–1635, 2005.

[34] L. Shi, H. Sun, X. Lu, C. Deng, and Y. Liu, "Study of the improved MM-Wave omni-directional microstrip antenna," in *Proceedings of the 7th International Symposium on Antennas, Propagation and EM Theory (ISAPE '06)*, pp. 238–241, October 2006.

[35] L. Shi, H. Sun, and X. Lu, "A composite antenna with wide circularly polarized beamwidth," *Microwave and Optical Technology Letters*, vol. 51, no. 10, pp. 2461–2464, 2009.

[36] L. Shi, H. Sun, W. Dong, and X. Lv, "A dual-band multi-function carborne hybrid antenna for satellite communication relay system," *Progress in Electromagnetics Research*, vol. 95, pp. 329–340, 2009.

[37] Y. Chen, H. Sun, and X. Lv, "Novel design of dual-polarization broad-band printed L-shaped probe fed microstrip patch antenna for SAR applications," in *IET International Radar Conference 2009*, pp. 1–4, April 2009.

[38] L. M. Si, H. J. Sun, and X. Lv, "Numerical simulations of backward-to-forward leaky-wave antenna with composite right/left-handed coplanar waveguide," *Chinese Physics Letters*, vol. 27, no. 3, Article ID 034106, 2010.

[39] L.-M. Si, H.-J. Sun, and X. Lv, "THz leaky-wave antenna with high-directivity and beam-steering using CPW CRLH meta-material resonators," in *Proceedings of the 3rd International Symposium on Photoelectronic Detection and Imaging*, Beijing, China, 2009.

[40] Y. Liu, L. M. Si, S. Zhu, and H. Xin, "Experimental realization of an integrated THz electromagnetic crys-tals (EMXT) H-plane horn antenna," *Electronics Letters*, vol. 47, no. 2, pp. 80–82, 2011.

[41] L. M. Si, Y. Liu, S. H. Zhu, and H. Xin, "Integrated THz horn antenna using EBG structures," in *Proceedings of the IEEE International Symposium on Antennas and Propagation and USNC/CNC/URSI National Radio Science Meeting*, University of Colorado at Boulder, 2011.

[42] Y. Liu, Y. Yuan, X. Lv, and L.-M. Si, "Design of 0.5 THz 2D square lattice EBG waveguide transmission line and power-divider using MEMS technology," in *Proceedings 3rd International Symposium on Photoelectronic Detection and Imaging*, vol. 7385, Beijing, China, 2009.

[43] L.-M. Si, Y. Yong, H.-J. Sun, and L. Xin, "Characterization and application of planar terahertz narrow bandpass filter with metamaterial resonators," in *Proceedings of the International*
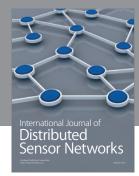
Workshop on Metamaterials, pp. 351–354, Nanjing, China, 2008.

[44] L. M. Si and X. Lv, "Terahertz waves hairpin microstrip band-pass filter and its application to overlaid dielectric material detection," *Modern Physics Letters B*, vol. 22, no. 29, pp. 2843–2848, 2008.

[45] L.-M. Si, H.-J. Sun, and X. Lv, "Theoretical investigation of terahertz amplifier by carbon nanotubes within transmission line metamaterials," *Microwave and Optical Technology Letters*, vol. 53, no. 4, pp. 815–818, 2011.

[46] W. Hong, "Millimeter wave and THz communications in China," in *Proceedings of the IEEE International Microwave Symposium Digest (MTT '11)*, pp. 1–4, 2011.
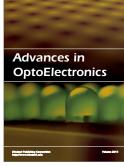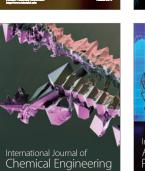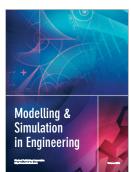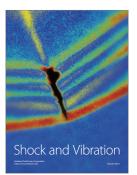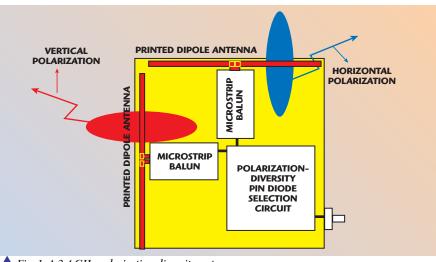
A 2.4 GHz Polarization-diversity Planar Printed Dipole Antenna for WLAN and Wireless Communication Applications

This article presents the design simulation, fabrication and measured performance of a 2.4 GHz polarization-diversity printed dipole antenna for wireless communication applications. Two orthogonal printed dipole antennas, each with a microstrip via-hole balun for vertical and horizontal polarization, are combined and fabricated on a PCB substrate. PIN diodes are used as switches to select the desired antenna polarization. The 3D finite-element-method (FEM) electromagnetic EM simulator, HFSS, is used in the design simulation of this planar antenna structure. Numerical and measured results of the antenna radiation characteristics, including input SWR, radiation pattern coverage and polarization-diversity, are presented and compared.

In wireless communication systems, such as wireless local area networks (WLAN), research and development efforts are aiming at smaller size and better performance. In addition to the use of signal processing techniques to improve communication channel capacity, the radiation characteristics of the portable antenna system is also very important for communication performance.

In urban or indoor environments, the radio wave will propagate through complicated reflection or scattering processes. The polarization of the radio wave may change significantly. In order to effectively receive the communications signal, a polarization-diversity antenna for wireless communications may become an important requirement. A polarization-diversity antenna may have a pair of linearly-polarized antennas, and the radio signal received on both antenna is sampled and compared at

HUEY-RU CHUANG, LIANG-CHEN KUO, CHI-CHANG LIN AND WEN-TZU CHEN
National Cheng Kung University, Tainan, Taiwan

▲ Fig. 1 A 2.4 GHz polarization-diversity antenna.
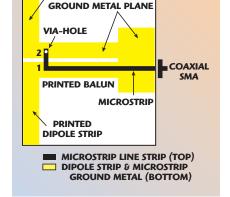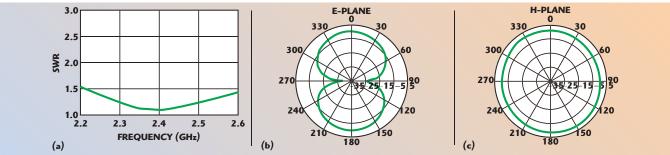
▲ Fig. 2 Printed dipole antenna with a microstrip balun.

▲ Fig. 3 Simulated performance of a 2.4 GHz printed dipole antenna placed horizontally; (a) input SWR, (b) E-plane pattern and (c) H-plane pattern.

certain time intervals. Then the antenna with the best signal quality is selected.

A typical dipole antenna radiates a vertically polarized EM wave and has an omnidirectional antenna pattern. In order to have a preferred planar antenna structure for this 2.4 GHz polarization-diversity antenna, a printed dipole antenna with a microstrip via-hole balun is designed. As shown in **Figure 1**, two orthogonal printed dipole antennas, for vertical and horizontal polarization, respectively, are combined and fabricated on a PCB substrate. PIN diodes are used to switch and select the desired antenna polarization.

In the antenna design, the high frequency structure simulator (HFSS), based on a 3D FEM, was employed for design simulation of the complete printed dipole structure. A printed dipole antenna and a polarization-diversity planar dipole antenna board (with a polarization-selection PIN diode circuit) have been fabricated on FR-4 PCB substrates. A complete 3D structure FEM simulation and the measured performance

of the realized printed dipole-antenna are compared. The measured radiation characteristics of the polarization-diversity planar dipole antenna, including input SWR, radiation pattern coverage and polarization diversity, are presented.

PRINTED DIPOLE ANTENNA WITH MICROSTRIP BALUN
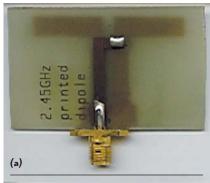
As shown in **Figure 2**, a printed dipole antenna has a printed microstrip balun which acts as an unbalanced-to-balanced transformer from the feed coaxial line to the two printed dipole strips. The length of the dipole strip and the balun microstrip are both about 1/4 wavelength. The ground plane of the microstrip line and the dipole antenna strips are in the same plane. A via-hole permits the feed point 2 of a printed dipole strip to have the same phase as the feed point 1 of the other printed dipole strip. Due to the 180° phase difference between the top strip and the ground plane of the microstrip line, the feed point 2 of the printed dipole strip will have 180° phase difference with the other feed point 1. Accurate

dimensions of the printed dipole strip and the microstrip balun structure are determined by numerical simulation, using HFSS.

The simulation results for a 2.4 GHz printed dipole antenna placed horizontally with a microstrip via-hole balun and fabricated on an FR-4 substrate are shown in **Figure 3**. The input SWR is less than 1.5 from 2.2 to 2.6 GHz. The simulated E- and H-plane antenna patterns are very close to those of an ideal dipole antenna, where the H-plane pattern is omnidirectional. **Figure 4** is a photograph of a realized antenna. The measured input SWR and antenna patterns (measured with the dipole placed vertically) agree well with the simulation results, as shown in **Figure 5**.

PLANAR POLARIZATION-DIVERSITY PRINTED DIPOLE ANTENNA

Figure 6 shows photographs of a realized 2.4 GHz planar polarization-diversity antenna consisting of two orthogonal printed dipole antennas with a polarization-switched PIN diode circuit. Each printed dipole has

▲ *Fig. 4 A 2.4 GHz printed dipole antenna with a microstrip via-hole balun; (a) top view and (b) bottom view.*

a microstrip via-hole balun. The terminals of the two baluns are connected to a PIN diode selection circuit. Voltages from the transceiver circuit (±5.0V) are fed through a cable to the input of the PIN diode circuit section, to short or open-circuit the PIN diodes. Hence, either the vertical or horizontal printed dipole can be selected and connected to the transceiver.

Since the two dipoles are very close to each other and near the PIN diode circuit section, EM coupling will degrade the performance of each dipole. **Figure 7** shows the input SWR simulation results with the vertical dipole antenna selected (+5V to PIN diode switching circuit). The input SWR is less than 2 from 2.25 to 2.60 GHz. The simulated E- and H-plane antenna patterns are all very close to those of an ideal dipole antenna, of which the H-plane pattern is still omnidirectional, as shown in **Figure 8**. Note that the dominant polarization is the vertical (E_θ) field, which agrees with the selection of the vertical dipole. The antenna pattern has some attenuation in the direction of the PIN diode circuit board. It can also be seen that a certain level of the input RF signal is induced to the horizontal antenna path by EM coupling, which generates some level of cross-
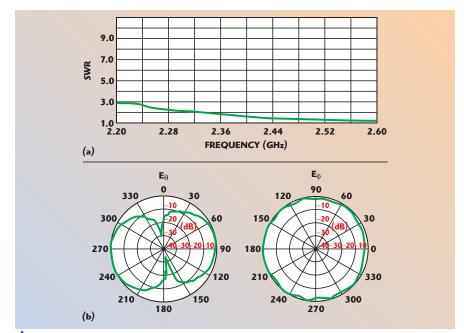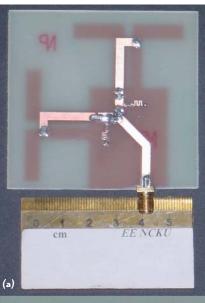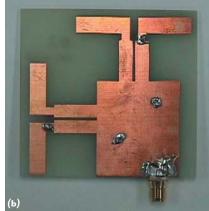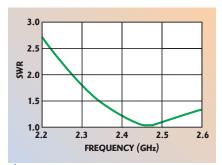
▲ *Fig. 5 Measured input SWR (a) and radiation patterns (b).*
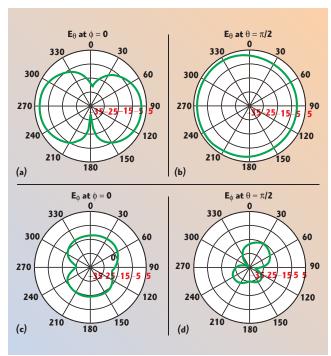
▲ *Fig. 6 A 2.4 GHz planar polarization-diversity antenna with a polarization-switched PIN diode circuit; (a) top view and (b) bottom view.*

▲ *Fig. 7 Input SWR simulation of a 2.5 GHz polarization-diversity dipole antenna with the vertical dipole selected.*

polarization field. **Figure 9** shows the simulation results with the horizontal dipole antenna selected (–5V to PIN diode switching circuit). Results similar to the ones obtained for the vertical dipole antenna can be observed, except that the dominant polarization is the horizontal (E_ϕ) field, which agrees with the selection of the horizontal dipole.

The measured antenna input SWR with vertical or horizontal dipole selection confirms the input SWR of each dipole antenna (through the PIN diode selection circuit) is less than 1.5 from 2.2 to 2.6 GHz, which agrees with the HFSS simulation results. The measured antenna patterns with the selection of the vertical or horizontal dipole shows that for the selection of the vertical dipole, the H-plane pattern is still quite omnidirectional (as an ideal vertical dipole) with some attenuation in the direc-
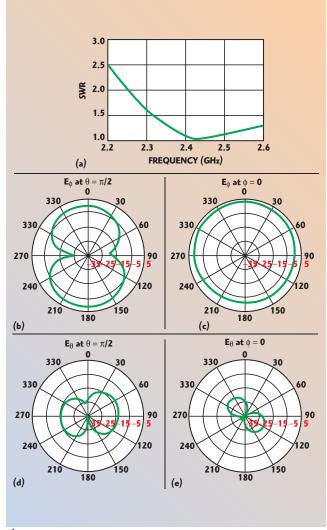
▲ *Fig. 8 Simulation of a 2.5 GHz polarization-diversity dipole antenna with the vertical dipole selected; (a) E_θ-field E-plane pattern, (b) E_θ-field H-plane pattern, (c) E_ϕ-field (cross-polarization) E-plane pattern and (d) E_ϕ-field (cross-polarization) H-plane pattern.*
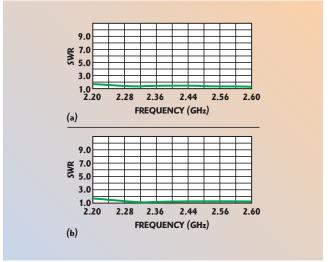
tion of the PIN diode circuit board. **Figures 10** and **11** show the mesured SWR and antenna patterns, respectively. A certain level of the induced cross-polarization pattern is observed as predicted by the HFSS simulation due to the proximity of the horizontal dipole strip and the PIN diode circuit board. As for the selection of the horizontal dipole, the E-plane pattern is also close to that of an ideal horizontal dipole. Also, the induced cross-polarization pattern is observed, which is the same situation as the selection of the vertical dipole. The measured data shows a good agreement with the HFSS simulation results and how the antenna polarization-diversity is working.

CONCLUSION
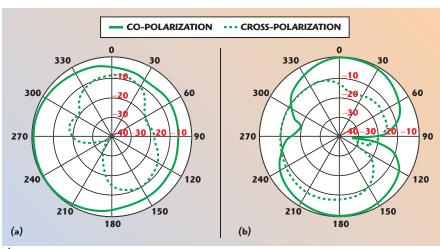
3D FEM design simulation, realization and measurements of a 2.4 GHz printed dipole antenna (with a microstrip via-hole balun) and a planar polarization-diversity printed dipole antenna are presented. The planar polarization-diversity antenna consists of two orthogonal printed dipole antennas for vertical and horizontal polarization and is fabricated on a FR-4 PCB board. A PIN diode switching circuit is used to select the desired antenna polarization. Satisfactory agreement between simulation and measurements is observed. The measured input SWR of the realized printed dipole antenna is less than 1.5 from 2.2 to 2.6 GHz. The measured input SWR of the vertical and horizontal dipole (through the PIN diode switching circuit) of the realized planar polarization-diversity antenna is less than 1.5 from 2.3 to 2.6 GHz. The measured E- and H-plane patterns of the polarization-diversity antenna show that the selected vertical or horizontal dipole have a performance close to a single dipole antenna in a vertical or horizontal position. The designed planar polarization-diversity antenna can be used for wireless communication and WLAN applications.

▲ *Fig. 9 Simulation of a 2.4 GHz polarization-diversity printed dipole antenna (with the horizontal dipole selected); (a) input SWR, (b) E_ϕ-field E-plane pattern, (c) E_ϕ-field H-plane pattern, (d) E_θ-field (cross-polarization) E-plane pattern and (e) E_θ-field (cross-polarization) H-plane pattern.*

▲ *Fig. 10 Measured input SWR of a 2.4 GHz polarization-diversity printed dipole antenna; (a) vertical dipole selection and (b) horizontal dipole section.*

TECHNICAL FEATURE

▲ *Fig. 11 Measured co-and cross-polarized patterns of a 2.4 GHz polarization-diversity printed dipole antenna; (a) vertical dipole selected and (b) horizontal dipole selected.*

ACKNOWLEDGMENT

References

1. K. Fujimoto and J. R. James, *Mobile Antenna Systems Handbook*, Artech House Inc., Norwood, MA 1994.
2. B. Edward and D. Rees, "A Broadband Printed Dipole with Integrated Balun," *Microwave Journal*, May 1987, pp. 339–344.
3. K. Hettak, G.Y. Delisle and M.G. Stubbs, "A Novel Variant of Dual Polarized CPW Fed Patch Antenna for Broadband Wireless Communications," *IEEE Antennas and Propagation Society International Symposium Digest*, Vol.1, 2000, pp. 286–289.
4. L. Zhu and K. Wu, "Model-based Characterization of CPS-fed Printed Dipole for Innovative Design of Uniplanar Integrated Antenna," *IEEE Microwave and Guided Wave Letters*, Vol. 9, No. 9, September 1999, pp. 342–344.
5. N. Michishitai and H. Arai, "A Polarization-diversity Antenna by Printed Dipole and Patch With a Hole," *IEEE Antennas and Propagation Society International Symposium Digest*, 2001, pp. 368–371.

Huey–Ru Chuang *received his BSEE and MSEE degrees from National Taiwan University, Taipei, Taiwan, in 1977 and 1980, respectively, and his PhD degree in electrical engineering from Michigan State University, East Lansing, MI, in 1987. From 1987 to 1988, he was a post-doctoral research associate at the Engineering Research Center of Michigan State University. From 1988 to 1990, he was with the Portable Communication Division of Motorola Inc., Ft. Lauderdale, FL. He joined the department of electrical engineering of National Cheng Kung University, Tainan, Taiwan, in 1991, where he is currently a professor. His research interests include portable antenna design, RF/microwave circuits and RFIC/MMIC for wireless communications, electromagnetic computation of human interaction with mobile antennas, EMI/EMC, microwave communication and detection systems.*

Liang–Chen Kuo *received his BSEE degree from Nan-Tai Institute of Technology, Tainan, Taiwan, and his MSEE degree from Tatung Institute of Technology, Taipei, Taiwan, in 1987 and 1996, respectively. He is currently working toward his PhD degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan. His research interests include computational electromagnetics and antenna design.*

Chi–Chang Lin *received his BSEE and MSEE degrees from Tatung Institute of Technology, Taipei, Taiwan, in 1999 and 2001, respectively. He is currently working toward his PhD degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan. His research interests include EM simulation and microwave antenna design.*

Wen–Tzu Chen *received his PhD degree from National Cheng Kung University, Tainan, Taiwan, in 1998. He is currently an assistant professor at the Institute of Computer and Communication, Shu-Te University, Yen Chau, Taiwan. His research interests include numerical computation of EM interaction between the antenna and the human body, and microwave antenna design.*

射 频 和 天 线 设 计 培 训 课 程 推 荐

　　易迪拓培训(www.edatop.com)由数名来自于研发第一线的资深工程师发起成立，致力并专注于微波、射频、天线设计研发人才的培养；我们于 2006 年整合合并微波 EDA 网(www.mweda.com)，现已发展成为国内最大的微波射频和天线设计人才培养基地，成功推出多套微波射频以及天线设计经典培训课程和 ADS、HFSS 等专业软件使用培训课程，广受客户好评；并先后与人民邮电出版社、电子工业出版社合作出版了多本专业图书，帮助数万名工程师提升了专业技术能力。客户遍布中兴通讯、研通高频、埃威航电、国人通信等多家国内知名公司，以及台湾工业技术研究院、永业科技、全一电子等多家台湾地区企业。

　　易迪拓培训推荐课程列表：　http://www.edatop.com/peixun/tuijian/

射频工程师养成培训课程套装

该套装精选了射频专业基础培训课程、射频仿真设计培训课程和射频电路测量培训课程三个类别共 30 门视频培训课程和 3 本图书教材；旨在引领学员全面学习一个射频工程师需要熟悉、理解和掌握的专业知识和研发设计能力。通过套装的学习，能够让学员完全达到和胜任一个合格的射频工程师的要求…

　　课程网址：http://www.edatop.com/peixun/rfe/110.html

手机天线设计培训视频课程

该套课程全面讲授了当前手机天线相关设计技术，内容涵盖了早期的外置螺旋手机天线设计，最常用的几种手机内置天线类型——如 monopole 天线、PIFA 天线、Loop 天线和 FICA 天线的设计，以及当前高端智能手机中较常用的金属边框和全金属外壳手机天线的设计；通过该套课程的学习，可以帮助您快速、全面、系统地学习、了解和掌握各种类型的手机天线设计，以及天线及其匹配电路的设计和调试...

　　课程网址：　http://www.edatop.com/peixun/antenna/133.html

WiFi 和蓝牙天线设计培训课程

该套课程是李明洋老师应邀给惠普 (HP)公司工程师讲授的 3 天员工内训课程录像，课程内容是李明洋老师十多年工作经验积累和总结，主要讲解了 WiFi 天线设计、HFSS 天线设计软件的使用，匹配电路设计调试、矢量网络分析仪的使用操作、WiFi 射频电路和 PCB Layout 知识，以及 EMC 问题的分析解决思路等内容。对于正在从事射频设计和天线设计领域工作的您，绝对值得拥有和学习！…

　　课程网址：http://www.edatop.com/peixun/antenna/134.html

CST 学习培训课程套装

该培训套装由易迪拓培训联合微波 EDA 网共同推出，是最全面、系统、专业的 CST 微波工作室培训课程套装，所有课程都由经验丰富的专家授课，视频教学，可以帮助您从零开始，全面系统地学习 CST 微波工作的各项功能及其在微波射频、天线设计等领域的设计应用。且购买该套装，还可超值赠送 3 个月免费学习答疑…

　　　　课程网址：http://www.edatop.com/peixun/cst/24.html

HFSS 学习培训课程套装

该套课程套装包含了本站全部 HFSS 培训课程，是迄今国内最全面、最专业的 HFSS 培训教程套装，可以帮助您从零开始，全面深入学习 HFSS 的各项功能和在多个方面的工程应用。购买套装，更可超值赠送 3 个月免费学习答疑，随时解答您学习过程中遇到的棘手问题，让您的 HFSS 学习更加轻松顺畅…

　　　　课程网址：http://www.edatop.com/peixun/hfss/11.html

ADS 学习培训课程套装

该套装是迄今国内最全面、最权威的 ADS 培训教程，共包含 10 门 ADS 学习培训课程。课程是由具有多年 ADS 使用经验的微波射频与通信系统设计领域资深专家讲解，并多结合设计实例，由浅入深、详细而又全面地讲解了 ADS 在微波射频电路设计、通信系统设计和电磁仿真设计方面的内容。能让您在最短的时间内学会使用 ADS，迅速提升个人技术能力，把 ADS 真正应用到实际研发工作中去，成为 ADS 设计专家...

　　　　课程网址：　http://www.edatop.com/peixun/ads/13.html
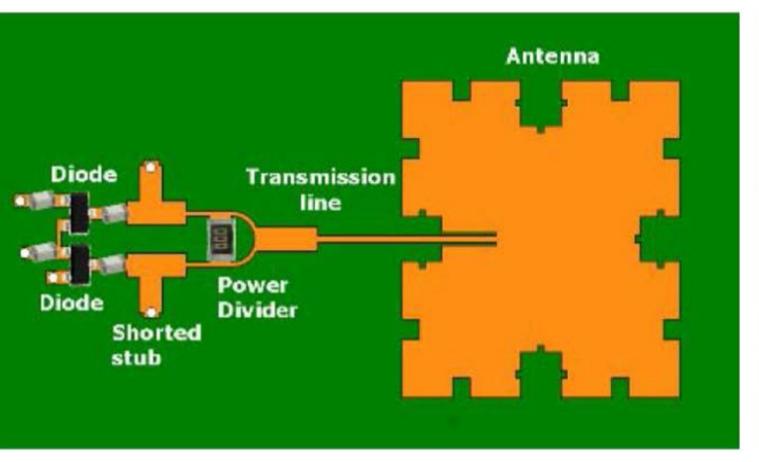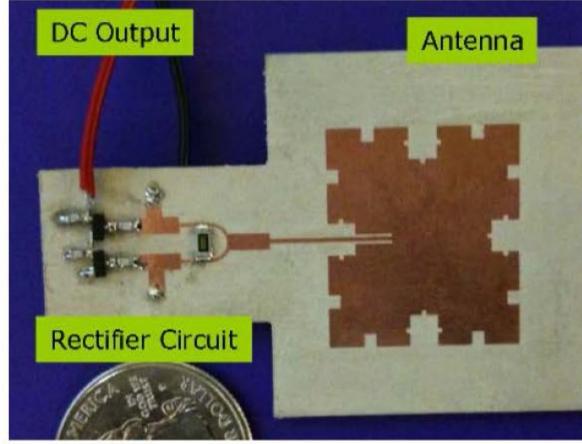
我们的课程优势：

※ 成立于 2004 年，10 多年丰富的行业经验，

※ 一直致力并专注于微波射频和天线设计工程师的培养，更了解该行业对人才的要求

※ 经验丰富的一线资深工程师讲授，结合实际工程案例，直观、实用、易学

联系我们：

※ 易迪拓培训官网：http://www.edatop.com

※ 微波 EDA 网：http://www.mweda.com

※ 官方淘宝店：http://shop36920890.taobao.com

Giedrius Rakauskas
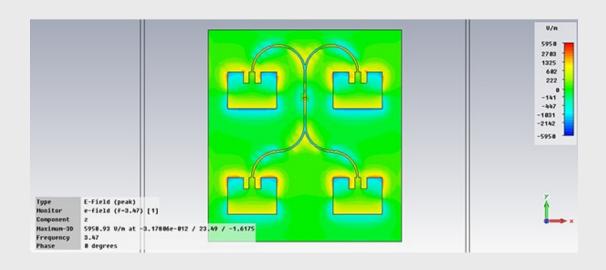
Phased Patch Antenna Array Design

Engineering
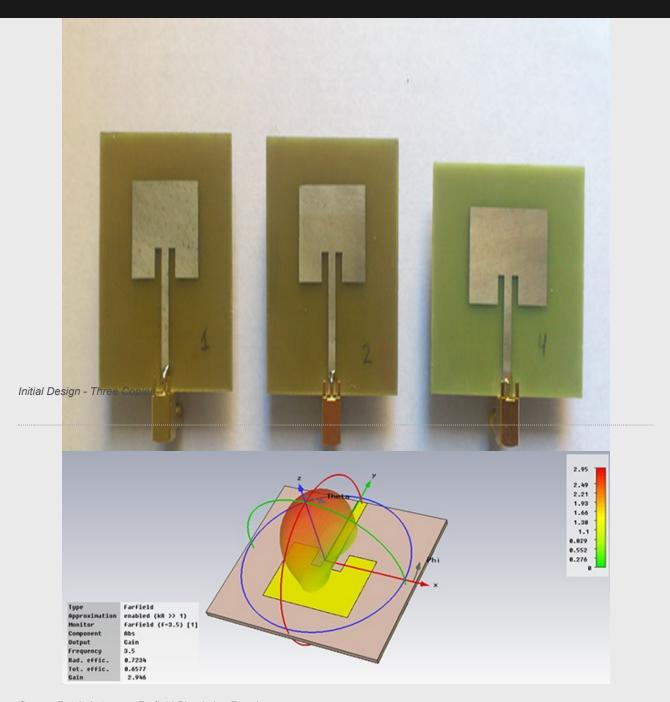
Third Year Project

The aim of this project was to design a directive phased patch antenna array that could be implemented in modern vehicles for distance measurement and autonomous cruise control.
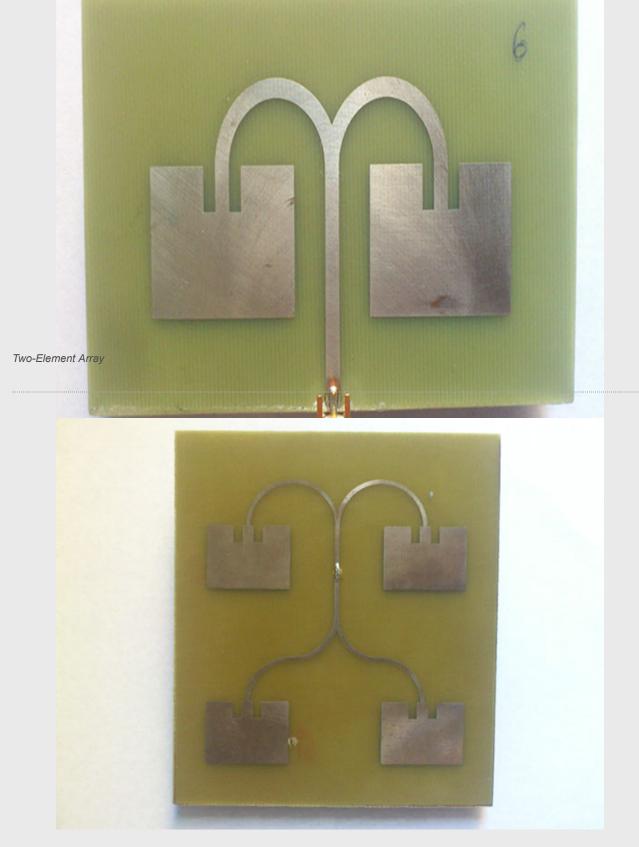
The particular technology allows improved performance as well as reduced costs and size.

Initial Design - Three Copies
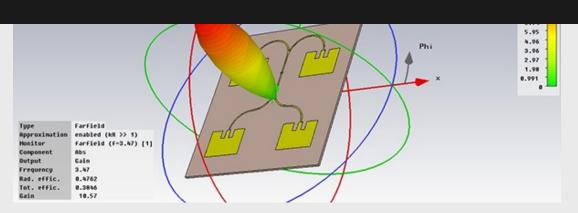
Square Patch Antenna - Farfield Simulation Results
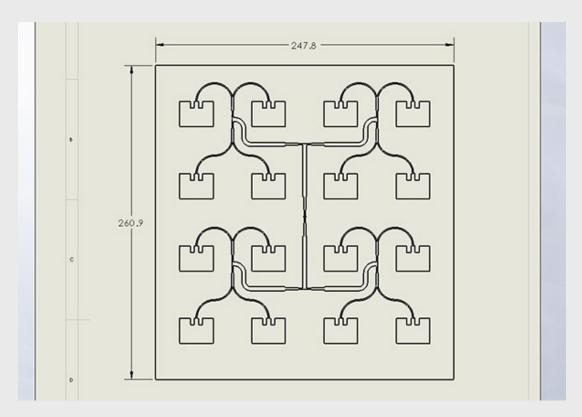
Two-Element Array

Four-Element Array

Type	Farfield
Approximation	enabled (kR >> 1)
Monitor	farfield (f=3.47) [1]
Component	Abs
Output	Gain
Frequency	3.47
Rad. effic.	0.4762
Tot. effic.	0.3846
Gain	10.57
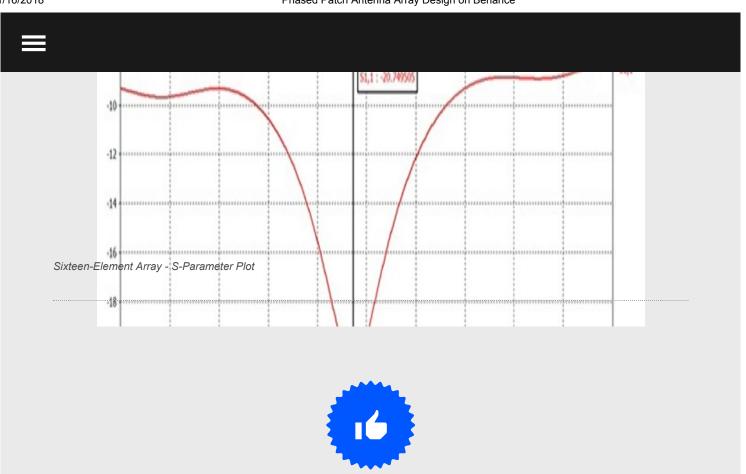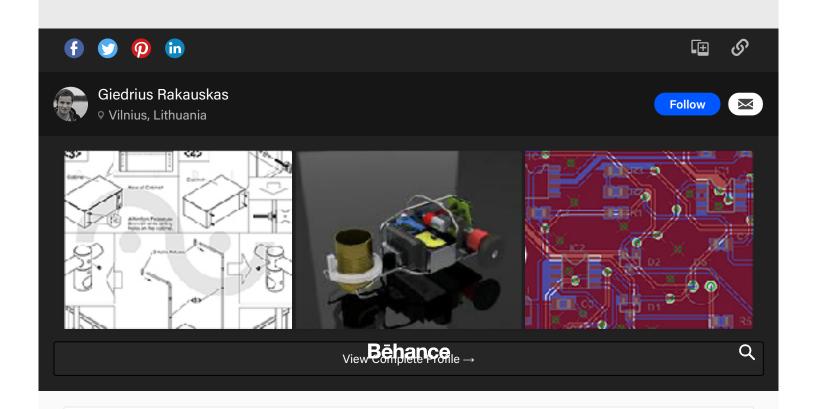
Four-Element Array - CST Microwave Studio Simulation Results

Sixteen-Element Array Design

S1,1 : -20.749595

Sixteen-Element Array - S-Parameter Plot

Giedrius Rakauskas
 Vilnius, Lithuania

Follow

Comments

Basic Info

Third Year BEng Project - Phased Patch Antenna Array

Credits

Giedrius Rakauskas
⦿ Vilnius, Lithuania

+ Follow

Tags

| patch antenna | antenna array | square patch antenna | cst mws | cst microwave studio |

Tools Used

CST Microwave Studio, FR-4

⚠ Report

Bēhance 🅰
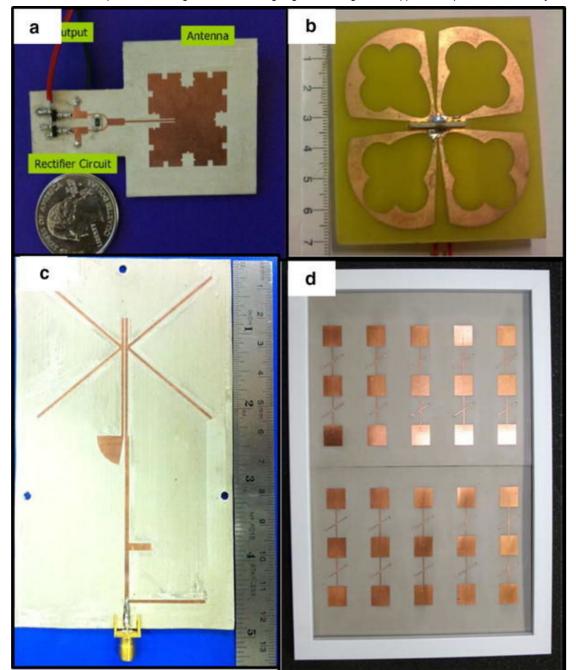
Fig. 4

Types of antennas **a** 2.45 GHz patch antenna with rectifier [38]. ©2010 IEEE. All rights reserved; **b** Planar dual-polarized antenna [24]. ©2015 IEEE. All rights reserved; **c** Microstrip antenna [37]. ©2012 IEEE. All rights reserved; **d** Array of stacked differential patch antenna [33]. ©2015 IEEE. All rights reserved

The plate antennas are popular and have many applications [27, 34, 38]; on-chip antennas are preferred for small and compact applications. Recently, many publications addressed wide-band and multi-band antennas. It has been proven that narrow-band antennas offer high energy conversion efficiencies but can only retrieve a limited amount of energy. On the other hand, wide-band or multi-band frequency antennas can retrieve more RF energy in space. However, the tradeoffs are low overall efficiency and large aperture. In [32], antennas with a resonance frequency of 4.9 and 5.9 GHz were designed with PCEs of 65.2 and 64.8%, respectively. Further work by Lu et al. [26] on polarization antennas supports the assertion that expanding the bandwidth of an antenna leads to increasing the amount of power harvested. In this work, the demonstration of broadband polarization antennas with three separate modes allows the antenna to operate in a wider range of frequencies. One common mechanism in the aforementioned works is the control of the antenna configuration by switching the diodes on and off, thus altering its resonant frequencies. However, since it uses separate modes for different frequencies, this antenna is not able to simultaneously resonate at two frequencies. On the other hand, the antenna presented in [39] is capable of operating at 2.45 and 5.8 GHz, simultaneously, providing 2.6 V output with a PCE of 65% and power density of 10 mW/cm^2.
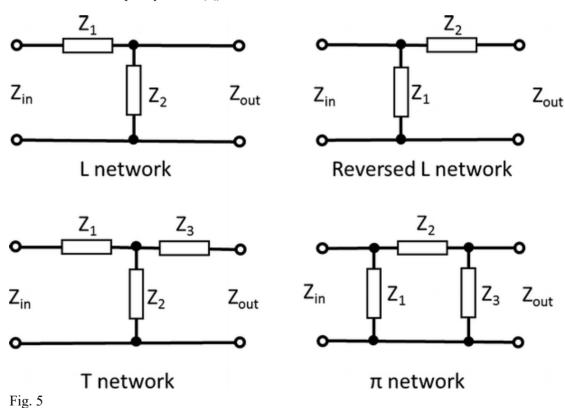
The main purpose of aligning antennas in arrays is to enhance the antenna gain and obtain high voltage/current. Array antennas are preferred over large aperture antennas because they do not require large breakdown voltage diodes to operate. Antenna arrays can be connected before or after rectification. The first configuration enhances the retrieved power at the main beam while the second configuration expands the ability to retrieve power from various angles away from the main beam [40]. In case the RF waves are combined before rectification, the rectifier requires a large breakdown diode. If RF waves are combined after rectification, combining DC current becomes problematic. Antenna arrays can be connected in series or parallel to obtain high voltage or large current. Nonetheless, expanding the arrays yields better output but this might cause a deduction in conversion efficiency [33].

For demonstrating antenna arrays, Sun et al. [35] invented a T-junction to connect four quasi-Yagi antennas together. The advancement in this work was that the T-junction was flexible in changing from 1×4 array to 2×2 array topologies. Consequently, the system was able to operate at an ambient power level as low as 455 μW/cm^2 while obtaining 40% PCE.

Impedance matching network

In low-power consumption electrical systems, power leakage during transmission may lead to energy insufficiency. In these circumstances, adding an impedance matching network (IMN) ensures that the maximum power transfers between the RF source and load. For WPH applications, the receiving antenna is considered as the source while the rectifier/voltage multiplier is considered as the load. It is acknowledged that in DC, power transfer is optimum when the resistances of the source and load are indistinguishable. In an RF circuit, the impedance is referred to instead of resistance. An impedance mismatch between the source and load creates reflected power flow in the circuit that lowers the efficiency of the system. As its name indicates, the IMN ensures that the impedance of the source and load are identical by adding reactive components in between.

There are three basics matching configurations i.e. L, T and π matching networks (Fig. 5). The L matching is commonly used since it typically has two components, which simplifies the designing and controlling process. Additionally, the L matching networks do not alter the quality factor (Q) of the circuit.

Fig. 5

Configuration of common impedance matching networks

The T and π matching configurations are more complex than the L network. Furthermore, organizing the T and π configurations into multiple stages will retain the final matching results but will change the Q factor. This strategy is useful in improving voltage boost.

There are tradeoffs between the attributes of an IMN, which include frequency, bandwidth, adjustability, and complexity. For instant, in [41], the suboptimal impedance matching and multiport ladder matching methods were introduced to enhance the

matching performance and harvested power of antennas, respectively. However, the tradeoff was that implementation of these configurations required more components than traditional matching networks, thus, escalating the circuit's complexity.
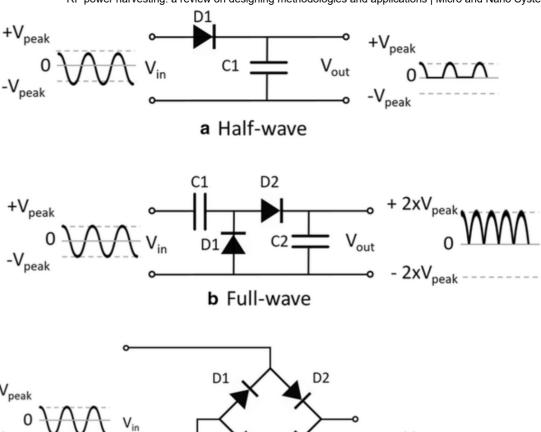
Etor et al. [42] designed an IMN for THz frequencies applications using transmission lines and self-designed metal–insulator-metal diodes instead of lumped components. Moreover, fixed IMN and tunable IMN [43, 44, 45] were introduced as a technique for better matching with wide-band and multi-band antennas.

Rectifier/voltage multiplier

RF energy extracted from free space usually possesses low power density since the electric field power density decreases at the rate of $1/d^2$, where d is the distance from the RF source [46]. Therefore, a power amplifier circuit is required that yields enough DC energy from the electromagnetic waves to drive the loads. This gives rise to two possibilities, if the power consumption of the load is lower than the average power harvesting, the electronic devices at the load may work continuously; otherwise, if the load consumes more energy than the power harvesting circuit can generate, the devices cannot work continuously [47].

Rectifying is the most popular application of diodes, which refers to the conversion of AC current to DC current. In terms of power harvesting application, the RF signal retrieved in the antenna has a sinusoidal waveform. The signal after transformation through IMN would be rectified and boosted to meet the power requirements of the application.

The most fundamental topology of the rectifier is the half-wave rectifier that comprises of a single diode D1 (Fig. 6a). When AC voltage transfers through D1, only the positive cycle remains and the negative cycle is cutoff; thus, it diminishes half of the AC power. Moreover, the output V_{out} is discontinuous since the negative cycle is cutoff. Despite its simplicity, a half-wave rectifier is usually inadequate for common applications. Hence, a full-wave rectifier is more preferable. The circuit design of the full-wave rectifier is shown in Fig. 6b. During the first negative cycle of AC input, diode D1 is conductive and capacitor C1 is charged to the corresponding energy level of V_{peak} of the input. Then, at the next positive cycle, diode D1 is blocked, diode D2 is conductive so that capacitor C2 is also charged. In consequence, the output V_{out} would see two capacitors in series (each one is storing a voltage of $V_{peak)}$. Thus, V_{out} is twice V_{peak}. Therefore, this topology is more stable and efficient than the half-wave rectifier. There is also a bridge rectifier that rectifies both positive and negative cycles of the AC input but retains $V_{out} = V_{peak}$ by alternatively blocking pairs of diodes D1, D4 and D2, D3 (Fig. 6c).
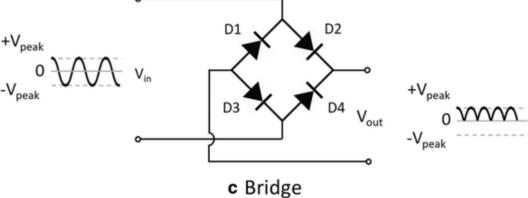
Fig. 6

Some common topologies of a rectifier

Voltage multiplier is a special type of rectifier circuit that converts and boosts AC input to DC output. In some case where the rectified power is inadequate for the application, there is a need for boosting the output DC by stacking single rectifiers into series, forming the voltage multiplier [48]. Several configurations of the voltage multiplier are shown in Fig. 7. The most fundamental configuration is the Cockcroft–Walton voltage multiplier (Fig. 7a). This circuit's operational principle is similar to the full-wave rectifier (Fig. 6b) but has more stages for higher voltage gain. The Dickson multiplier in Fig. 7b is a modification of Cockcroft–Walton's configuration with stage capacitors being shunted to reduce parasitic effects. Thus, the Dickson multiplier is preferable for small voltage applications. However, it is challenging to obtain high PCE due to the high threshold voltage among diodes creating leakage current, thus reducing the overall efficiency. Additionally, for high resistance loads, output voltage drops drastically leading to low current supply to the load. A summary of recent works related to voltage multiplier is shown in Table 3.
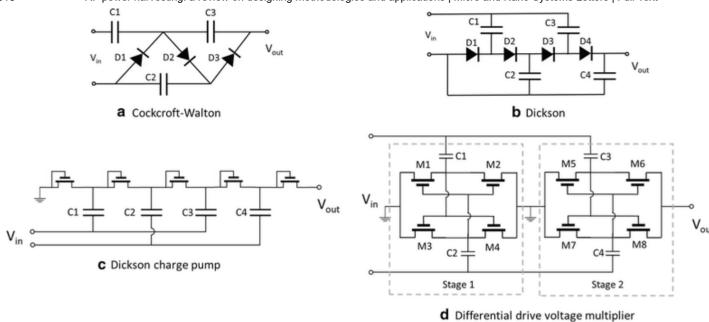
Fig. 7

Common voltage multiplier configurations: **a** Three stages Cockcroft–Walton voltage multiplier, **b** Four stages Dickson voltage multiplier, **c** Four stage Dickson voltage multiplier using CMOS technology, **d** Two stages voltage multiplier comprised of differential drive unit

Table 3

Published work comparison regarding voltage multiplier

Ref.	Rectifier topology	No. of stage	Freq. (MHz)	Tech.	Range (dBm)	Maximum PCE	Size
[51]	Differential	2	433	0.18 μm CMOS	0–20	74% @ 2 dBm	–
[52]	PMOS transistors	7	900	40 nm CMOS	–	44% @ > 390 mV	0.04 mm^2
[53]	Half-wave	4	900	0.18 μm CMOS	–	37.42% @ 390 mV	–
[54]	Comparator-based/active-diode	3	13.56	0.18 μm CMOS	8–15	67.9% @ 12.8 dBm	–
[55]	Dickson	3	13.56	250 nm CMOS	–	72%	0.13 mm^2
[56]	–	–	915	0.13 μm CMOS	−21.6	22.6% @ −16.8 dBm	0.186 mm^2

Prototyping Methods

There's more on this in Scherz, *Practical Electronics for Inventors.*

Solderless Breadboard or Plugboard

This is what we've been using in lab.

Advantages

- Very fast to build and make changes.
- Works well with DIP ICs

Disadvantages

- Limited reliability—lab testing only, for limited-size circuits.
- High capacitance between adjacent rows (~ 10 pF).
- Only for small-lead-size components (.032" or 0.82 mm max—a 1 A diode lead is just barely too big to meet the spec.)

Springboard

Used in ENGS 22

Advantages
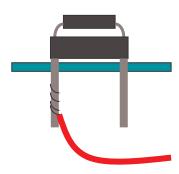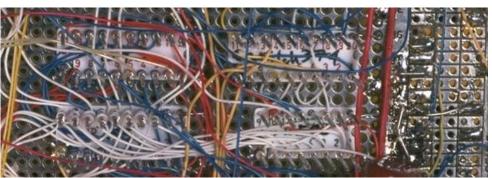
- Fast to build and make changes.
- Accomodates large wire sizes.
- Secure, reliable connections.
- Can handle higher current/power.

Disadvantages

- Can't accommodate small lead spacings such as on ICs.
- Although it is reliable enough for long-term use, it's expensive for that purpose.

Wirewrap

Connections made by fine wire wrapped tightly around square pins of special IC sockets. This is a great way to make a permanent version of a digital circuit almost as quickly as using a solderless breadboard. Used in ENGS 31.

Advantages

- Fast to build and make changes.
- Can make a complex circuit compact.
- Reliable.
- Can be used permanently.
- Inexpensive.

Disadvantages

- Low current—limited to under a few hundred mA.
- High inductance/resistance—on the order of 0.1 ohm, 0.25 µH per foot.
- Only makes reliable connections to square leads (as are on wire-wrap IC sockets). Discrete components (e.g., resistors, capacitors, transistors) need to be soldered to a "header" that goes in an IC socket or to individual wire-wrap pins.

Perfboard and Solder.

Using the same perfboard as used in wire-wrap work, it is possible to simply twist and solder leads, and run wires where needed.

Advantages
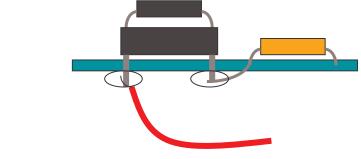
- Reliable, if done well.
- Can be used permanently.
- Inexpensive.
- Can handle any size components
- Convenient for working with discrete components.

Disadvantages
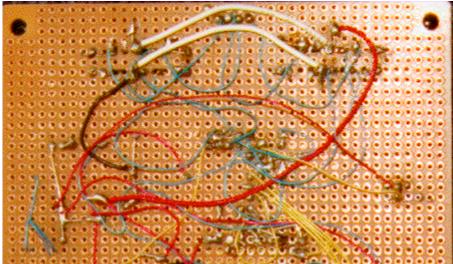
o Slow, requires skill to do well.
o Works for ICs, but not very easily.

Variations:
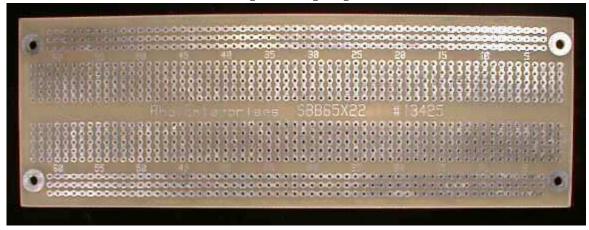
- Perfboard with individual copper pads on each hole so that the solder will better hold things in place. (Photo below)
- Perfboard with a perforated ground plane. Provides the shielding and grounding benefits of a ground plane, and makes ground connections easy, but requires care to avoid shorts. Some have the ground plane etched from around the holes; others require you to cut the copper away from the holes with a special tool. This is also possible with wire-wrap, as in the photo in that section.
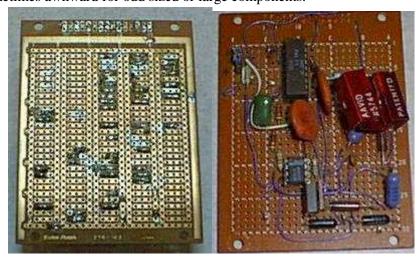
Generic Printed-Circuit Board—multiple hole-per-pad.
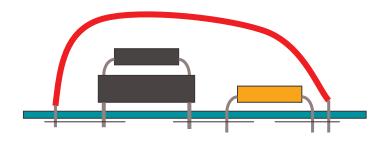
A printed circuit board with a pattern of holes and connections similar to a solderless breadboard.

Advantages

- Easier than using plain perfboard, especially for ICs.
- Reliable.
- Can be used permanently.
- Available with ground planes if needed.

Disadvantages

o Usually not as compact a final circuit as some alternatives, because you are constrained by layout.
o Pads bigger than needed can add capacitance, but not much.
o Can be expensive, especially the "vectorboard" brand.
o Sometimes awkward for odd sized or large components.

"Dead Bug," or "Ugly-board"

Start with a plain copper-clad board. Glue ICs down with the leads sticking up in the air. Then solder to them

Advantages	Disadvantages
• Provides an excellent ground plane. Can be a high performance way to build sensitive and/or high-frequency analog circuits	o Requires high soldering skill.
	o Takes a long time to build.
• Can make a complex circuit compact.	o Mechanical support for components is marginal; can add glue ("RTV") after debugging.
• Reliable.	o Only makes reliable connections to square leads (as are on wire-wrap IC sockets). Other components (e.g., resistors, capacitors, transistors) need to be soldered to a "header" that goes in an IC socket or to individual wire-wrap pins.
• Can be used permanently.	
• Inexpensive.	
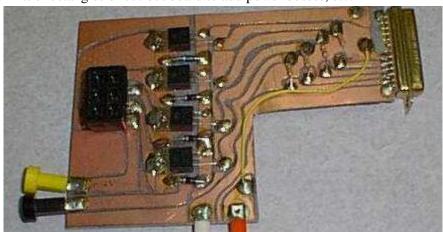
Variations:
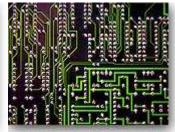
- Manually cut the board with a dremel tool to isolate sections for purposes other than ground plane (use the back for ground plane). See photo below.
- Glue on little rectangles of cut-out board to add power busses, etc.

Printed circuit board (PCB)
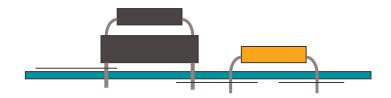
This is what is used virtually universally in production of electronics.

Advantages
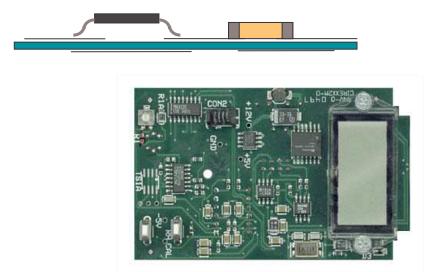
- Easy to build in production.
- Repeatable, controllable stray L, C.
- Can handle virtually any component, power level.
- Highly reliable.
- Can make very compact.
- Design can be (somewhat) automated from a schematic you have entered.

Disadvantages

o Laying out the board and getting it fabricated takes time, although you can pay for fabrication in a few days if you can afford it.

o Expensive, on the order of hundreds of dollars for one, but with almost no increase in cost to make many.

o Hard to make changes, but making changes may be easier than building another type of prototype.
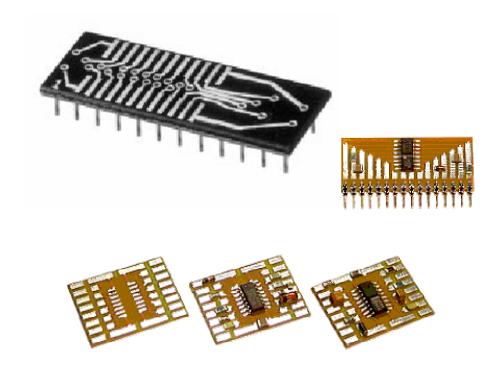
Notes on surface-mount components

Most modern production designs now use surface-mount components instead of "through hole" components. The circuits can be more compact, and the board layout is easier because different things can be done on each side (and in additional layers between sides) without through-holes interfering. But prototyping gets much more difficult!

Prototyping options for surface mount

- Simulate, and then lay out a PCB. Don't ever make a breadboard.
- Order DIP ICs and leaded passives for the prototype, and then switch to surface mount for production.
- Get adaptors that have pads to solder surface-mount ICs to, and then standard-spacing (0.1") pins in DIP or SIP layout. Digikey carries "surfboards" made by Capital Advanced Technologies, and adaptors made by Aries Electronics. In addition to DIP and SIP adaptors, there are boards with solder pads for connecting larger wires; these work well for prototypes built in "dead bug" style.

Crystal receiver set 11

Back to the index

This receiver has a single tuned circuit, the Q factor of the tuned circuit is quite low.
The receiver can be used well for reception of local stations, for reception of distant stations it is less suitable.

I tried to give this receiver a nice "old-fashioned" look, the reception performance was in this design of less importance.
This receiver is for sale, I made a series production of these, and will build more on request.
This receiver is also for sale as do-it-yourself kit.
See the shop for more information.
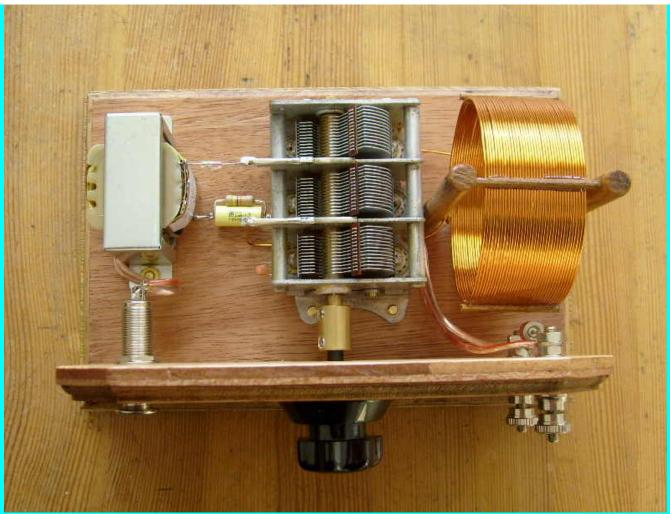
The front panel of the receiver.
On the left side the socket for the headphone.
In the middle the tuning knob.
On the right the connections for antenna (upper) and ground (lower).
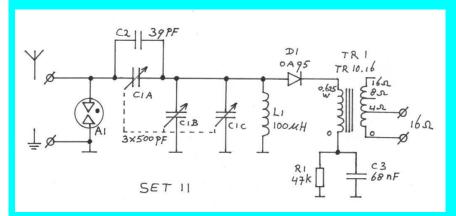
Top view of the receiver.

Circuit diagram of receiver set 11

Circuit description.
The resonance circuit is formed by coil L1 and C1b and C1c (together 1000 pF).
Capacitor C1a and C2 provide the matching between antenna and tuned circuit.
The frame (rotor) of the tuning capacitor is carrying the RF signal, by this it is possible to tune simultaneously the tuned circuit (C1b and C1c) and the antenna matching (C1a).
Germanium diode D1 provides the signal detection.
Transformer TR1 is loaded with 16 Ω at its 4 Ω output, through this the input impedance increases from 16 kΩ to about 43 kΩ.
At the output a headphone of 2x 32 Ω can be connected, with the two speakers parallel, the impedance is 16 Ω.

Component A1 is a *gas discharge tube* (also called: *surge arrester*) with type number: N81-A90X.
The gas discharge tube protects the antenna input for too high voltages.

These high voltages can occur if the antenna picks up static charge from the air (especially occurs with long outdoor antennas from non insulated antenna wire).
As the voltage at the antenna is higher then 90 Volt, the gas discharge tube will start to conduct and short the high voltage to ground.
As soon as the charge has flown to ground, the conduction stops automatically.

Frequency range of the receiver.
Frequency range without antenna: 500 - 2500 kHz.
Frequency range with 10 meter antenna connected: 487 - 1860 kHz.
Both with and without antenna connected, the whole medium wave band can be tuned.

Q factor of the LC circuit (Without antenna and without diode connected):
600 kHz: Q= 83
900 kHz: Q= 81
1200 kHz: Q= 75
1500 kHz: Q= 65
The circuit Q is rather low, one reason for this is because the RF signal is in this design on the frame of the tuning capacitor.
Because the frame of the tuning capacitor is directly connected to the wooden bottom plate losses occur here.

You can find a complete building instruction op the following pages:
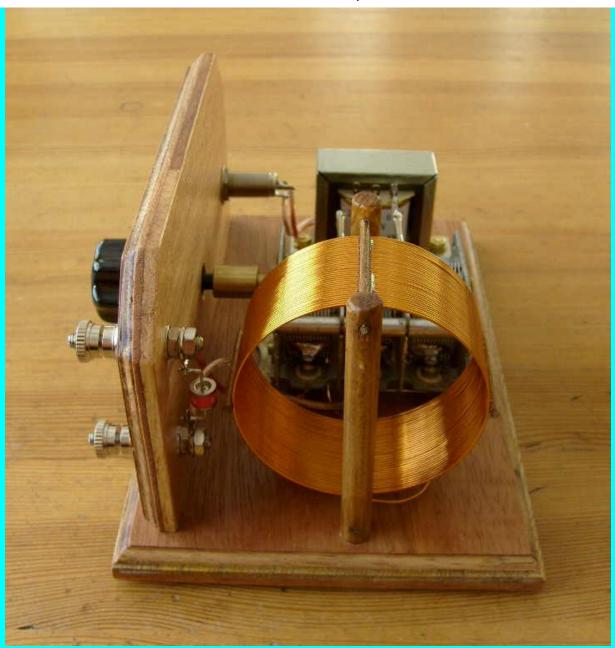
Step 1 Preparing the components
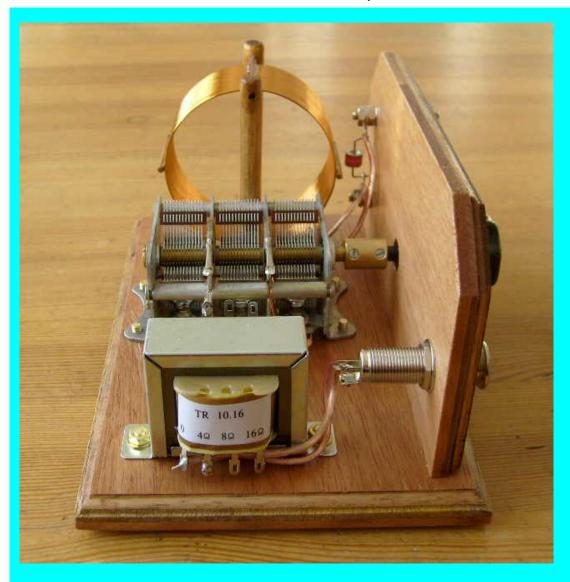
Step 2 Making the frame of the receiver

Step 3 Making the coil

Step 4 Placing the components

 You will find here the part list of this receiver.
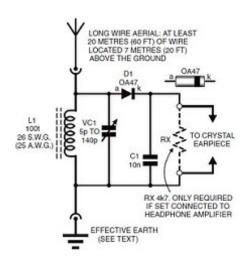
[Back to the index](http://www.crystal-radio.eu/set11/enset11.htm)

How to make a batteryless (crystal set) radio

By **Sagar Sapkota** - July 21, 2012

CRYSTAL SET

The simplest radio receiver, known as a Crystal Set, consists of nothing more than a coil, tuning capacitor, diode detector, and a pair of earphones. A typical circuit diagram for a Crystal Set Radio is given below where inductor or coil L1 is tuned by variable capacitor VC1 to the transmitter frequency. Diode D1 demodulates the signal, which is fed straight to the earphones. There is no amplification.

[CLICK HERE FOR BIGGER PHOTO](#)

You can use 1N34 Germanium diode in place of OA47

A long (at least 20 metres), high (17 metres or more) aerial and a good earth (a buried biscuit tin or a metre of copper pipe driven into damp ground) are required in order to ensure audible headphone reception. The earphones originally used with these receivers had an impedance of around 4000 ohms and were very sensitive (and heavy and uncomfortable). They are no longer available, but a crystal earpiece, which relies on the piezoelectric effect, will give acceptable results.

Low impedance ″Walkman″ type earphones are NOT suitable.

Component details:

Resistor- RX-4.7k, 0.25 W- only required if set is connect to audio amplifier

Capacitors- C1- 10nF disc ceramic VC1- 5p to 140p, polythene dielectric variable capacitor.

Semiconductors: D1: OA47 or 1N34- Germanium Diode

Miscellaneous: L1- Ferrite Rod, 100mm(4 inch)x9mm/10mm dia., with coil.

Crystal earpiece and jack socket to suit; plastic control knob; plastic insulated flexible cable for aerial wire, downlead and earth connection, 30 meters minimum; buried biscuit tin or 1 meter of copper pipe for earth system; 50gm reel of 26SWG enamelled copper wire, for tuning coil; card and glue for coil former; multistrand connecting wire; crocodile clips or terminals for aerial and earth lead connection; solder, etc.

READ MY EXPERIENCES- HOW EASILY I MADE A BATTERYLESS RADIO

DRAWBACKS

Quite apart from the absence of amplification, two factors seriously limit the performance of crystal receivers. Germanium diodes become increasingly reluctant to conduct as the applied voltage falls below 0.2V, and this makes the receiver insensitive to weak signals. Silicon diodes have a threshold of around 0.6V, and are, therefore, unsuitable for circuits of this kind.

Silicon diodes like 1N4001 are not suitable
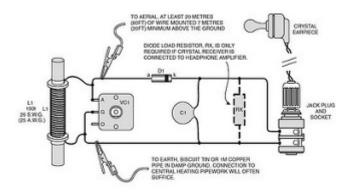
The earphone loading imposes heavy damping on the tuned circuit, hence, reduces its ability to separate signals. With such low selectivity insensitivity can be a blessing, and crystal sets are normally only capable of receiving a single, strong transmission on the long and medium wavebands. They will sometimes receive more than one if a shortwave coil is fitted.

The aerial and diode can be connected to tappings on the tuning coil in order to reduce damping, but the improvement in selectivity is usually at the expense of audio output. When valves cost a week's wages and had to be powered by large dry batteries and lead/acid accumulators, the☐ construction of simple receivers of this kind could be justified. With high performance transistors now costing only a few pence or cents, crystal sets are now regarded as 'nostalgic pieces". Some readers may, however, wish to build one out of curiosity, or for the novelty of having a receiver that **does not require a power supply.**
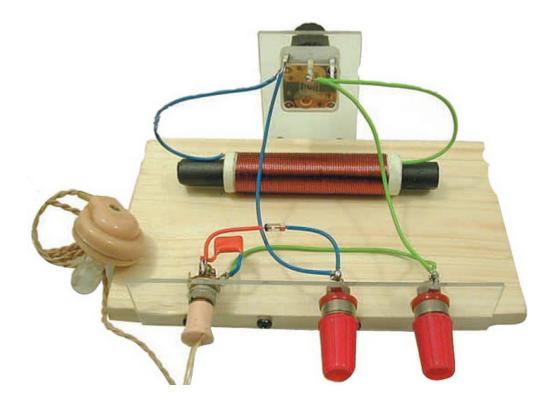
CIRCUIT DETAILS

Ferrite loop aerial L1 and polythene dielectric variable capacitor VC1 form the tuned circuit. Point contact germanium diode D1 (1N34 or OA47) demodulates the signal; capacitor C1 bypasses residual r.f. (radio frequency) to earth and also exhibits a reservoir action, enabling the a.f. (audio frequency) output to approach its peak value. The recovered audio signal is fed directly to a crystal earpiece. Signal voltages

introduced in the ferrite loop aerial by the radiated magnetic field are much too much to produce an output from the detector, and the component is used here simply as a tuning coil. The ferrite core does, however, reduce the number of turns required for the coil winding, thereby reducing its resistance and increasing its audio quality.
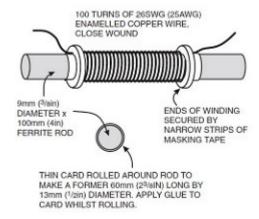
GET BIGGER PICTURE

COIL DETAILS

Full construction and winding details for the ferrite tuning/ aerial coil L1 are shown in figure.

The coil is made from 26 SWG(25 AWG) enamelled copper wire, close wound on a cardboard former.

The r.f. bypass capacitor C1 can in practice, be omitted with no noticeable reduction in performance. However, if the set is to be connected to either the headphone amplifier or speaker amplifier decribed next month, the component together with diode, load resistor, RX, must be included.

AMPLIFICATION Audio frequency amplification, after the diode detector will permit the use of low impedance Walkman type earphones or even loud speaker operation. It will do nothing, however, to overcome the diode's insensitivity to weak signals. For this we must have radio frequency amplification of the signals picked up by the aerial before they reach the detector(The standard circuit for a transistor portable receiver has three stages of radio frequency amplification ahead of the diode).

For amplification, you can use a simple audio amplifier using LM386.

READ MY EXPERIENCES ON MAKING A BATTERYLESS RADIO

WATCH THESE VIDEOS:

- **JLCPCB: Prototype 10 PCBs for $2 (2-layer,100*100mm)**
- **China's Largest PCB Prototype Manufacturer, 290,000+ Customers & 8,000+ Online Orders Per Day**
- **PCB Pricing: https://jlcpcb.com/quote**

Fully printed flexible and disposable wireless cyclic voltammetry tag

11 authors, including:

Jinsoo Noh
Gumi Electronics and Information Technolog…
22 PUBLICATIONS   641 CITATIONS

SEE PROFILE

Myoungho Pyo
Sunchon National University
96 PUBLICATIONS   1,319 CITATIONS

SEE PROFILE

Ali Javey
University of California, Berkeley
296 PUBLICATIONS   22,530 CITATIONS

SEE PROFILE

Gyoujin Cho
Sunchon National University
95 PUBLICATIONS   2,833 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Electrochemistry View project

Fully printed flexible and disposable wireless cyclic voltammetry tag

Younsu Jung[1], Hyejin Park[1], Jin-Ah Park[1], Jinsoo Noh[1], Yunchang Choi[1], Minhoon Jung[2], Kyunghwan Jung[2], Myungho Pyo[1], Kevin Chen[3], Ali Javey[3] & Gyoujin Cho[1]

[1]Department of Printed Electronics Engineering, Sunchon National University, Maegok, Sunchon, Jeonnam 540-742 Korea, [2]Research Institute of Printed Electronics, PARU Co. Seomyeon, Sunchon, Jeonnam 540-813 Korea, [3]Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720.

A disposable cyclic voltammetry (CV) tag is printed on a plastic film by integrating wireless power transmitter, polarized triangle wave generator, electrochemical cell and signage through a scalable gravure printing method. By proximity of 13.56 MHz RF reader, the printed CV tag generates 320 mHz of triangular sweep wave from +500 mV to −500 mV which enable to scan a printed electrochemical cell in the CV tag. By simply dropping any specimen solution on the electrochemical cell in the CV tag, the presence of solutes in the solution can be detected and shown on the signage of the CV tag in five sec. 10 mM of N,N,N′,N′-tetramethyl-p-phenylenediamine (TMPD) was used as a standard solute to prove the working concept of fully printed disposable wireless CV tag. Within five seconds, we can wirelessly diagnose the presence of TMPD in the solution using the CV tag in the proximity of the 13.56 MHz RF reader. This fully printed and wirelessly operated flexible CV tag is the first of its kind and marks the path for the utilization of inexpensive and disposable wireless electrochemical sensor systems for initial diagnose hazardous chemicals and biological molecules to improve public hygiene and health.

C yclic voltammetry (CV) has been used as a powerful tool for the study of electrochemical redox reactions between the electrodes and the solutions of all kind of solutes such as metals[1], organic molecules[2], proteins[3], bacteria[4], viruses[5], and DNA[6]. Because of the high sensitivity of the electron transfer redox reactions, CV could become a very promising ubiquitous electrochemical sensor protocol if the cost and size of the CV system were significantly reduced to commercially viable single-use disposable units for checking traces of hazardous materials such as lead[7], mercury[8], arsenic[9], e-*coli*[10], and pesticides[11] in water or to diagnose the level of glucose[12], cholesterol[13] and specific enzymes[14] in blood. This single-use disposable CV would dramatically reduce the cost of maintaining the public health system[15]. Therefore, inexpensive and disposable CV measurement system that can be operated wirelessly using a smartphone or RF (radio frequency) reader without complicate operation processes is in high demand for the realization of an ubiquitous sensor network system (Figure 1a). They would mainly be used as ubiquitous diagnostic and testing tools for detecting and monitoring the level of target specimens. However, there is no technology that is advanced enough yet to build a wireless, inexpensive and disposable CV system. In this paper, as a form of RF-tag, an extremely inexpensive, disposable and fully printed CV system is demonstrated for the first time by mimicking and combining basic CV concepts and wireless power transmission technologies of RF devices. To realize the fully printed CV tag, a key issue of wirelessly generating triangular waveform (±500 mV) that can scan the electrochemical cell at a low frequency (<1 Hz) to set up a redox reaction needs to be addressed through a minimum number of printed thin film transistors (TFTs). Fully printed 13.56 MHz rectenna and a ring oscillator with large trap charges in the channels of printed TFTs were respectively utilized to wirelessly generate triangular waveform by using only 10 printed TFTs.

Results
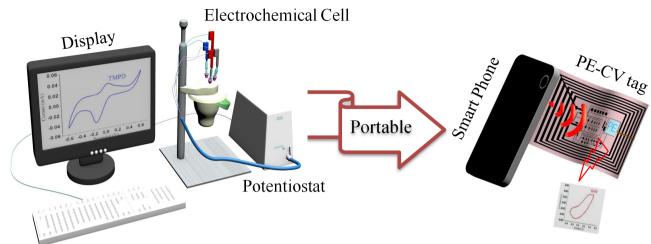
Based on two key units of the wireless power transmitter (Figure 1b-①) and the triangular wave generator (Figure 1b-②), the circuit layout of the fully printed wireless CV tag was designed by using a minimum number of printed thin film transistors (TFTs) to alleviate the issue of V_{th} shift and shown in Figure 1b as a platform of disposable CV system. The circuit of CV tag was fabricated using all scalable printing methods, a roll-to-roll gravure, a roll-to-plate gravure, a drop casting, and a screen printer (Figure S1 in Supplementary Information).

To provide a polarized DC voltage from the coupled 13.56 MHz AC signal of the reader, we modified our previously reported R2R gravure printed rectenna[16,17], as shown in the layout in Figure 2a and the resulting

a

b
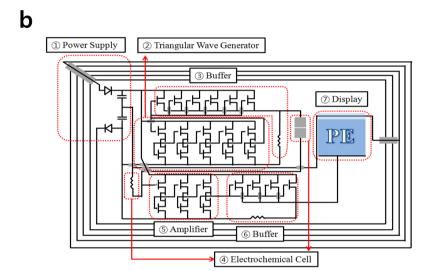
Figure 1 | (a) Informative illustration of typical CV system and disposable printed CV tag. (b) Schematic circuit diagram of the gravure printed wireless cyclic voltammetry (CV) tags. The circuit was designed to couple AC power from a 13.56 MHz reader and then convert the coupled AC to polarized DC ±10 V (① in Figure 1b). Polarized DC will operate the printed ring oscillator to generate a triangular voltage waveform (② in Figure 1b). The generated waveform will pass through the buffer to meet the impedance difference from the electrochemical cell (③ in Figure 1b). The electrochemical cell will run the redox reaction with a single drop of specimen solution (④ in Figure 1b) by the voltage triangular waveform. The output current of the electrochemical redox reaction will be amplified *via* an amplifier circuit (⑤ in Figure b) and the signal will pass through the buffer to meet the impedance difference from the printed signage (⑥ in Figure b). The signage will indicate the concentration level of specimen in the solution (⑦ in Figure b). It will indicate whether the concentration is above or below a pre-determined value.

printed diodes and capacitors are shown in Figure 2b. The printed diodes showed a rectifying ratio of 10^3, and capacitances of two printed capacitors were all about 8 nF/cm². The polarized DC voltages (Figure 2c) were measured by placing the printed rectenna on the RF (13.56 MHz) reader with a distance of 2 cm. We adapted a center tap transformer, consisted by divided antenna, 2 diodes and 2 capacitors, to provide the + and − DC voltage for printed triangle wave generator. Using a load of 1 MΩ, voltages of +9.4 V and −10.8 V DC were attained for the optimized printed antenna by monitoring the rectified polarized DC voltages and varying the values of the inductance of the antenna (Figure S2 in Supplementary Information). The polarized DC voltage was also decreased with decreasing load's resistance (Figure 2d).

To generate a cyclic waveform with a low frequency (<1 Hz) from the rectified polarized DC voltage, fully gravure printed *p*-type SWNT-based channel network thin film transistors (*cn*TFTs) were

used to construct a five stage ring oscillator with *p*-type inverters. The *cn*TFTs can operate under less than 20 V and their mobility range (0.01 to 10 cm²/Vs) can be controlled through the loading concentration of SWNT in the ink[18–21]. To print the five stage ring oscillator, we used the same silver and BaTiO₃ inks as in the printing of the rectenna for reducing a number of printing steps while SWNT ink was used for printing the channel of the *cn*TFTs. To manufacture the ring oscillator, roll-to-plate (R2P) gravure was used to print 10 *cn*TFTs with a gate width of 300 μm and channel length of 160 μm for both the drive and load *cn*TFTs (Figure S3 in Supplementary Information). The optimized cell depth of gravure plate was 37–40 μm with the cell wall thickness of 50–55 μm for the printing active layers of drive and load TFTs using low viscosity of SWNT based ink (~30 cp) (Figure S4 in Supplementary Information). By repeating the printing of active layers on load TFTs, a higher SWNT network density can be achieved to generate higher current in the

a

b

c

d
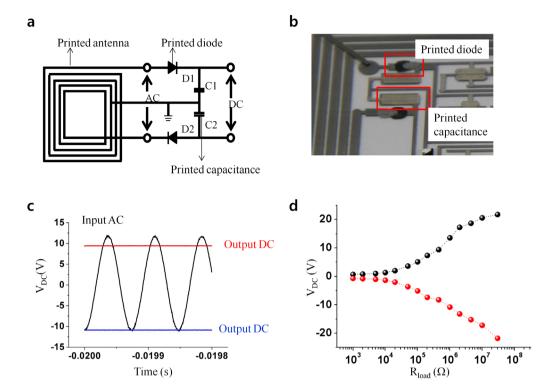
Figure 2 | (a) Circuit layout of the printed rectenna to provide polarized DC power, (b) a real image of printed diodes and capacitors, (c) input-output electrical characteristics of the rectifier at 13.56 MHz AC input and (d) output ± DC voltages under variations in load resistance.

channels while thicknesses of all printed gate, dielectric and drain-source electrodes are kept same (Figure S5 in Supplementary Information).

Their output and transfer characteristics are shown in Figure 3a and 3b respectively. Based on the attained transfer characteristics, the mobility (μ), transconductance (g_m), on-off current ratios from I_{DS}-V_{DS} and threshold voltages (V_{th}) for both five drive cnTFTs and five load cnTFTs were extracted and shown in Table S1 (Supplementary Information). The shift range of V_{th} from the fully gravure printed 10 cnTFTs were about ±0.4 V for the drive cnTFTs and ±0.6 V for load cnTFTs. Those shift ranges are narrow enough to run the printed ring oscillator for generating the triangular wave. Each inverter delivered a gain of 3 (Figure 3c), and the ring oscillator with five inverters generated a pseudo triangular waveform with a voltage amplitude of 7 V at 320 mHz (Figure 3d). In fact, the pseudo triangular wave can be only generated by the printed cnTFTs based ring oscillator because of the parasitic and trapped channel capacitances due to carbon nanotube network structures in cnTFTs. Although we only showed the frequency of 320 mHz at here, the frequency can be varied from 0.3 Hz to 2 Hz based on the printed network density of SWNT of printed cnTFTs. This is an advantage of our printed cnTFTs for scanning electrochemical cell in CV tag over other technologies such as amorphous silicon TFTs where only a sine wave with a designed single frequency can be generated because of no trap capacitance in the channel (Figure S6 in Supplementary Information). The frequency of the triangular waveform can be fine-tuned in the range of 0.3 Hz to 2 Hz as shown in the Figure S7 (Supplementary Information) by simply changing loading concentration of SWNTs.

When the generated triangular waveform is supplied directly into the printed electro-chemical cell (Figure S8 in Supplementary Information), wherein printed silver and carbon electrodes were used with a poly(ethylene oxide) and LiCF$_3$SO$_3$ gel type electrolyte, the output voltage range is reduced to zero because of improper impedance matching between the electrochemical cell (10 KΩ) and ring oscillator (1 MΩ). Therefore, a buffer unit is needed to match

the impedance levels between the ring oscillator and electrochemical cell. The printed buffer unit (Figure 1b-③) with high on currents and low on-off current ratios (Figure 3e) is consists of 6 cnTFTs and a resistor (~7 KΩ). In the printed CV tag, three resistors were printed by screen printer using a carbon paste (DC-20, purchased from Dozen Co. Korea) for the buffer units (Figure 1b-③ and ⑥) and the electrochemical cell (Figure 1b-④). Their electrical parameters are listed in Table S1 (Supplementary Information). The 6 cnTFTs of buffer unit were printed using R2P gravure with high SWNT concentration in the semiconducting ink. Each cnTFTs exhibited on-currents in the range of 400 ~ 500 μA with a very low on-off ratio due to the enhanced metallic percolation in the dense SWNT networks. The buffer units with low on-off ratio were employed to reduce the input voltage to ±500 mV (Figure 3f) which is the appropriate scanning range of the electrochemical cell. The printed electrochemical cell with two electrodes was proven to be well operated under commercial CV (Figure S9 in Supplementary Information). The resulting current after scanning printed electrochemical cell in the CV tag can be converted to voltage by connecting the printed resistor (Figure 1b-④: ~5 × 10³ ohm) which provides the output voltage for electrochromic indicator and re-plotted into the cyclic voltammogram.

Because these output voltage (±70 mV) and current (~40 μA) were low after scanning the electrochemical cell, they need to be amplified to turn on the electrochromic indicator for showing predetermined concentration (10 mM) level of N,N,N′,N′-tetramethyl-p-phenylenediamine (TMPD, purchased from Aldrich) as a standard reference in this work (Figure 3g). After running the redox reaction of TMPD (Figure 3h) in the printed cell, the output voltage and current needed to be amplified to reach 1.5 V and 500 μA, respectively, to provide sufficient power to display the letters "PE" as an indicator of the presence of TMPD in the specimen. To construct the amplifier, cnTFTs based three inverters were printed with a gain of three to five (Figure S10 in Supplementary Information) to amplify the output signal as shown in Figure 3i. The electrical characteristics of the 6 cnTFTs in the inverters with extracted mobility, transcon-
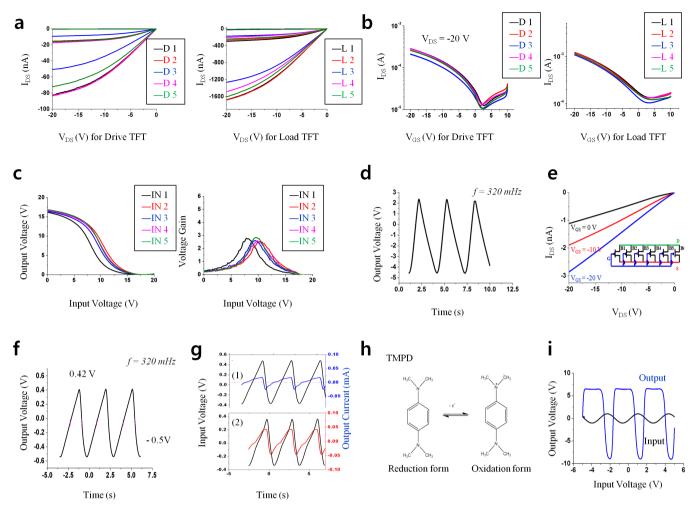
Figure 3 | (a and b) Output and transfer characteristics of the printed *cn*TFTs for 5 drive and 5 load TFTs respectively in the printed ring oscillator. (c) Electrical characteristics of inverters and (d) output characteristic of the ring oscillator. (e) Total output characteristics of the printed buffer unit consisting of 6 *cn*TFTs and a resistor (measured based on contacting gate and drain-source electrodes as shown in the inset circuit). (f) Modified triangular wave following the buffer unit. (g) Generated signals before scanning (black) and after scanning the electrochemical cells without (1, blue) and with TMPD (2, red) in a drop of solution. (h) TMPD structure for oxidation and reduction reaction. (i) The input and amplified output signals after passing through three amplifying inverters.

ductance, on-off current ratios and threshold voltage are shown in Table S2 (Supplementary Information). As the impedance was matched between the ring oscillator and the electrochemical cell using the buffer unit, another buffer unit (Figure 1b-⑥) is also needed between the electrochromic signage (20 KΩ) and the amplifier (1 MΩ) for impedance matching. The electrical parameters of the buffer are listed in Table S2 (Supplementary Information).

The amplified output signal was used to run the reduction and oxidation of the patterned conducting polymer in the electrochromic signage, and thus the blinking signage will be used to indicate the presence of chemicals in the specimens. The signage was fully printed and attached on printed CV tag. A clear concept of the printing sequence for the electrochromic signage was given in Figure S11 in Supplementary Information.

Discussion

The fully printed flexible CV tag was completed by assembling the printed circuits including ring oscillator, buffer, electrochemical cell, amplifier, and signage onto the previously R2R gravure printed rectenna. The resultant CV tag is shown in Figure 4a. The working concept of the wireless and flexible CV tag is demonstrated in the following sequences (watch the video file by clicking Figure S12 in Supplementary Information). After dropping 500 µl of TMPD solu-

tion (10 mM) on the printed electrochemical cell, the CV tag was placed on the custom made RF (13.56 MHz) reader (Figure S13 in Supplementary Information). The antenna was subsequently coupled to 13.56 MHz AC. The coupled AC was rectified into polarized DC ($> \pm 10$ V) through two diodes and two capacitors which caused the ring oscillator to generate a pseudo triangular waveform with output voltage of 7 V at 320 mHz. This was then passed through the buffer unit to scan the electrochemical cell. The output signal after scanning the electrochemical cell was amplified to turn on the signage according to the concentration level of TMPD in the solution. In this work, whenever the concentration of TMPD was higher than 10 mM, the signage "PE" blinked (Figure 4b) while it did not show anything (Figure 4c) when it was lower than 10 mM. Furthermore, clear cyclic voltammograms for scanning the electrochemical cell with and without 10 mM of TMPD were obtained by re-plotting the output triangular voltage waveform (Figure 4d). The attained half potential ($E_{1/2} \sim 0.05$ V, oxidation potential is 0.22 V and reduction potential is -0.17 V) of TMPD from the CV tag was nearly identical to the value obtained from a commercial CV instrument, SP-240, Biologic, which uses the same frequency for the triangular voltage waveform and printed electrochemical cell (Figure 4e). However, when the generated triangle wave frequency is higher than 0.6 Hz in the CV tag, the clear redox peaks cannot be
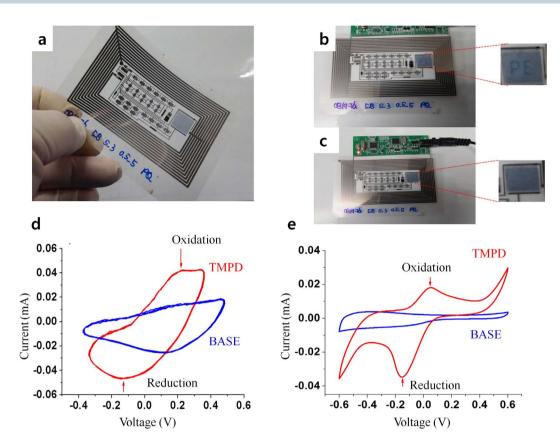
Figure 4 | (a) Optical image of fully printed wireless cyclic voltammetry tag. (b) Operation images of CV tag with and (c) without TMPD in the solution on 13.56 MHz reader. (d) Converted cyclic voltammogram from the printed wireless CV tag *vs* (e) a commercial CV instrument (please refer the interconnected video file for the demonstration of wireless CV tag operation in Figure S12 in Supplementary Information).

observed as shown in Figure S14 (Supplementary Information). Furthermore, as a typical example of testing a specimen in aqueous solution, 10 mM of $K_3(FeCN)_6$ aqueous solution was checked using the CV tag, and results of converted cyclic voltammogram was shown in Figure S15 (Supplementary Information). Those results support that the CV tag can examine specimens not only in organic solution, but also in aqueous one.

Up to present, there has been no commercial success in fully printed TFT-based electronic devices due to some kind of difficulties in integrating a number of printed TFTs through a scalable printing method. As increasing a number of integrated TFTs, the range of threshold voltage shits of TFTs needs to be controlled as narrow as possible to properly operate designed function of printed devices. However, the scalable printing method is not fully understood to integrate a number of TFTs on plastic foils while keeping the constant range of threshold voltage shifts because of difficulty in controlling a constant charge transport through the printed channels. Therefore, the printed TFT-based devices should be constructed by using a minimum number of printed TFTs while a function of printed device needs to be competitive over Si based one. Since the fully printed CV tag required the integration of only 26 *cn*TFTs, the acceptable range of threshold voltage shits of *cn*TFTs is controllable and achievable by using a scalable printing method such as a roll-to-roll gravure. Furthermore, the unique function of printed CV tags cannot be simply reproduced using a Si technology with a comparable cost to the R2R printing process. Therefore, the fully printed CV tag will be the first leading model to show a way of how printed TFT-based electronic devices should go for the successful launching in the IT markets.

In the core of the printed CV tag, fully gravure-printed *cn*TFTs were employed for constructing a circuit consisting of a ring oscillator to generate triangular waveforms, buffer units for an impedance

matching, and an amplifier for enhancing output signal to turn on the signage. Those simple units can be combined with a variety of printable sensors and RF devices to create novel devices with an extremely low cost. In other words, the printed rectenna, ring oscillator, buffer and amplifier units in the printed CV tag will be key elements for the construction of variety of printed RF-sensors, this CV tag will be used as a platform for the further construction of a variety of RF based electrochemical sensors.

In conclusion, we have presented a fully printed flexible CV tag with the size of 9.5×5.5 cm² that can operate wirelessly using a 13.56 MHz RF reader. However, for the practical purpose, the size of the CV tag can be reduced up to the half (4×2 cm²) of the current one by optimizing spaces between *cn*TFTs in the circuit of the CV tag and reducing the operation power to less than DC 5 V as the current size of 13.56 MHz antenna is designed to provide maximum coupled AC power using a sufficient length of printed antenna pattern (higher inductance). Furthermore, although we utilized R2P gravure, screen printer and drop casting method for the convenience of fabricating the CV tag in the Lab, the whole fabricating process can be repeated by R2R gravure because the rectenna, ring oscillator, buffers, electrochemical cell, amplifier and signage in the CV tag were all designed to compromise the limit (± 20 μm) of a registration accuracy of a current R2R gravure system. Therefore, this technology could be used as a platform for disposable wireless electrochemical sensors to a variety of specimens in aqueous and organic solutions to monitor or diagnose pathogens and hazardous materials by modifying the electrodes of the electrochemical cells.

Methods

Roll-to-roll (R2R) gravure printed antenna, bottom electrodes, gate electrodes and wires. Two color units of R2R gravure system (Taejin Co. Korea: Figure S1 in the Supplementary Information) was employed to print antenna, bottom electrodes and wires on a roll of poly(ethylene terephthalate) (PET) film (width of 200 mm and

thickness of 100 μm, purchased from SKC, Korea) with silver nanoparticle based conducting ink (PG-007 BB type, Paru Co. Korea) and the web transfer speed of 8 m/min under roll pressure of 0.5 MPa and web tension of 60 N. The silver ink was further formulated to meet the viscosity of 500 cp and surface tension of 47 mN/m using respectively ethylene glycol (Aldrich) and Dipropylene glycol methyl ether (Aldrich). The curing time of printed silver layer was 10 sec by passing through a heating chamber of 150°C, installed in R2R gravure system. The resulting printed silver patterns on PET roll showed the thickness of 450 (±50) nm without any defects and were shown in Figure S16a (Supplementary Information).

Roll-to-plate (R2P) gravure printed ring oscillator, buffer, and amplifier. Full R2P gravure system (Figure S1 in the Supplementary Information) was employed to print single walled carbon nanotube network based thin film transistors (*cn*TFTs), used as a building block to construct all the integrated circuit of printed cyclic voltammetry (CV) tag (Figure 1). 26 *cn*TFTs for the integration of circuit in printed CV tag were printed on the previously R2P gravure printed film (printed antenna, bottom electrodes and wires with thickness of 450 nm; Figure S16b in Supplementary Information) following printing sequences. First, $BaTiO_3$ nanoparticle based dielectric ink (PD-100, Paru Co., Korea) was further formulated to meet viscosity (75 cP) and surface tension (31.8 mN/m) using the ethyl 2-cynoacrylate (Aldrich). As shown in R2P gravure printing scheme (Figure S17 in Supplementary Information), the homogeneous and defect free dielectric layers were first R2P gravure printed with a printing speed of 200 mm/s and roll pressure 7.5 kg_f on previously printed gate electrodes and bottom electrodes (for capacitors) with a thickness of 2.3 (±0.1) μm and width of 940 (±2) μm (Figure S5 in Supplementary Information). The printed dielectric layers were cured for 1 min 20 sec at 150°C. Second, single walled carbon nanotube (SWNT) based semiconducting ink (PR-040, Paru Co., Korea) was diluted by diethylene glycol monobutyl ether (Daejung Co., Korea) to 1 : 1 volume ratio for printing active layers using R2P gravure on the printed dielectric layers with a printing speed of 350 mm/s and a roll pressure 4 kg_f. The viscosity and surface tension of diluted SWNT ink were respectively 30 cP and 29 mN/m. The resulting SWNT printed films were dried 1 min at 150°C. For using *cn*TFTs in the buffer unit, we used the dilution volume ratio of 5 : 1 between SWNT ink (PR-040, Paru Co., Korea) and diethylene glycol monobutyl ether for printing only on the TFTs in the buffer unit to increase SWNT network density for providing the high current (Figure S5 in Supplementary Information). Finally, drain-source electrodes were R2P gravure printed using silver nanoparticle based ink (PG-007 AA type, Paru Co., Korea) on SWNT printed film with a viscosity of 800 cP and a surface tension of 42 mN/m. The R2P gravure printing was carried out under a printing speed of 200 mm/s and a role pressure of 5 kg_f.

Diode fabrication. Although Shottky diodes were able to print using R2R gravure based on our previous reported process[16,17], the drop casting method was used for the convenience in this work. First, dropping ZnO based hybrid semiconducting ink (PD-070, Paru Co., Korea) on printed silver electrode and then, Al ink (PA-009, Paru Co., Korea) was dropped on the top of ZnO ink. The resulting diodes were further cured for 5 min at 120°C.

Measurements. Formulated and diluted inks were characterized using DCAT21 (Dataphysics Co., Germany) and SV-10 Vibro viscometer (AND Co., Japan) for respectively measure surface tension and viscosity. All measurements in this work were carried out under ambient condition if there are no other comments. Resistance of printed antenna, bottom electrodes, gate electrodes and wires were measured using two probe methods with HP3457A multimeter. Both surface morphology and thickness of printed layers was characterized using surface profiler (NV-2200, Nanosystem, Korea). LCR meter (E4980A, Agilent, USA), semiconductor parameter analyzer (4155C, Agilent, USA) and digital phosphor oscilloscope (DPO 4054, Tektronix, USA) were respectively used to characterize printed *cn*TFT based circuits.

1. Sanderson, M. D., Kamplain, J. K. & Bielawski, C. W. Quinone-annulated N-heterocyclic carbene-transition-metal complexes: Observation of π-back bonding using FT-IR spectroscopy and cyclic voltammetry. *J. Am. Chem. Soc* **128**, 16514–16515 (2006).
2. Lai, R. Y., Fabrizio, E. F., Lu, L., Jenekhe, S. A. & Bard, A. J. Synthesis, cyclic voltammetrics studies, and electrogenerated chemiluminescence of a new donor-acceptor molecule: 3,7-[Bis[4-phenyl-2-quinolyl]]-10-methylphenothiazine. *J. Am. Chem. Soc* **123**, 9112–9118 (2001).
3. Armstrong, F. A., Hill, H. A. & Walton, N. J. Direct electrochemistry of redox proteins. *Acc. Chem. Res* **21**, 407–413 (1998).
4. Norouzi, P., Ganjali, M. R., Daneshgar, P., Alizadeh, T. & Mohammadi, A. Development of fast Fourier transformation continuous cyclic voltammetry as a highly sensitive detection system for ultra-trace monitoring of penicillin V. *Anal. Biochem* **360**. 175–184 (2007).
5. Koyama, J. *et al.* Correlation of redox potentials and inhibitory effects on Epstein-Barr virus activation of 2-azaanthraquinones. *Cancer Letters* **212**, 1–6 (2004).
6. Teijeiro, C., Perez, P., Marin, D. & Paleček, E. Cyclic voltammetry of mitomycin C and DNA. *Bioelectrochemistry and Bioenergetics* **38**, 77–83 (1995).
7. Zerihun, T. & Gründler, P. Electrically heated cylindrical microelectrodes. Determination of lead on Pt by cyclic voltammetry and cathodic stripping analysis. *J. Electroanal Chem* **415**, 85–88 (1996).
8. Bond, A. M. & Scholz, F. Electrochemical, Thermodynamic, and mechanistic data derived from voltammetric studies on insoluble metallocenes, mercury halide and sulfide compounds, mixed silver halide crystals, and other metal complexes following their mechanical transfer to a graphite electrode. *Langmuir* **7**, 3197–3204 (1991).
9. Brusciotti, F. & Duby, P. Cyclic voltammetry study of arsenic in acidic solutions. *Electrochim Acta* **52**, 6644–6649 (2007).
10. Ruan, C., Yang, L. & Li, Y. Immunobiosensor chips for detection of escherichia coli O157:H7 using electrochemical impedance spectroscopy. *Anal. Chem* **74**, 4814–4820 (2002).
11. Liu, G. & Lin, Y. Electrochmical sensor for orgnohposphate pesticides and nerve agents using zirconia nanoparticles as selective sorbents. *Anal. Chem* **77**, 5894–5901 (2005).
12. Ascaso, J. F. *et al.* Diagnosing insulin resistance by simple quantitative methods in subjects with normal glucose metabolism. *Diabetes Care* **26**, 12, 3320–3325 (2003)
13. Karl, T. *et al.* Human breath isoprene and its relation to blood cholesterol levels: new measurements and modeling. *J. Appl. Physiol* **91**, 762–770 (2001).
14. Armstrong, F. A. Recent developments in dynamic electrochemical studies of adsorbed enzymes and their active sites. *Chem. Biol* **9**, 110–117 (2005).
15. Mollah, M. Y., Schennach, R., Parga, J. R. & Cocke, D. L. Electrocoagulation (EC)-science and applications. *J. Hazard. Mater* **B84**, 29–41 (2001).
16. Park, H. *et al.* Fully roll-to-roll gravure printed rectenna on plastic foils for wireless power transmission at 13.56 MHz. *Nanotech* **23**, 344006–344012 (2012).
17. Kang, H. *et al.* Fully Roll-to-Roll Gravure Printable Wireless (13.56 MHz) Sensor-Signage Tags for Smart Packaging. *Sci. Rep.* **4**, 5387; DOI: 10.1038/srep05387 (2014).
18. Jung, M. *et al.* All-printed and roll-to-roll-printable 13.56 MHz-operated 1-bit RF tag on plastic foils. *IEEE Trans Electron Devices* **57**, 3, 571–580 (2010).
19. Noh, J. *et al.* Integrable single walled carbon nanotube (SWNT) network based thin film transistors using roll-to-roll gravure and inkjet. *Org. Electron* **12**, 2185–5191 (2011).
20. Noh, J. *et al.* Modeling of printed single walled carbon nanotube thin film transistors for attaining optimized clock signals. *J. Appl. Phys* **108**, 10, 102811–102817 (2010).
21. Noh, J. *et al.* Fully Gravure-Printed Flexible Full Adder Using SWNT-Based TFTs. *IEEE. Eletron. Dev. Lett.* **33**, 11, 1574–1576 (2012).

Author contributions

Y.J., A.J. and G.C. designed the experiments. Y.J., H.P., J.P., Y.C. and M.J. carried out the experiments. J.N., K.J. and K.C. performed device simulations. Y.J. and K.C. performed mobility calculations. Y.J. and M.P. performed electrochemistry and analyzed electrochemical data. Y.J., H.P., M.P., A.J. and G.C. contributed to analyzing the data. M.J., M.P., A.J. and G.C. wrote the paper while all authors provided feedback.

Additional information

A microwave metamaterial with integrated power harvesting functionality

Allen M. Hawkes, Alexander R. Katko, and Steven A. Cummer

A microwave metamaterial with integrated power harvesting functionality

Allen M. Hawkes,[a] Alexander R. Katko, and Steven A. Cummer
Department of Electrical and Computer Engineering, Duke University, 101 Science Drive, Durham, North Carolina 27708-9976, USA

We present the design and experimental implementation of a power harvesting metamaterial. A maximum of 36.8% of the incident power from a 900 MHz signal is experimentally rectified by an array of metamaterial unit cells. We demonstrate that the maximum harvested power occurs for a resistive load close to 70 Ω in both simulation and experiment. The power harvesting metamaterial is an example of a functional metamaterial that may be suitable for a wide variety of applications that require power delivery to any active components integrated into the metamaterial. © 2013 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4824473]

Metamaterials are composed of sub-wavelength particles that exhibit bulk properties that are different from their individual components. Electromagnetic metamaterials are engineered materials that can achieve parameters not possible within naturally occurring materials, such as a negative index of refraction[1] or a zero index of refraction.[2] Exotic properties like these allow for a variety of interesting applications including a superlens device[3] and an invisibility cloak.[4] Integrating active and nonlinear functionality into metamaterials has been demonstrated in the form of dynamic resonant frequency tuning,[5,6] phase conjugation,[7] and wave mixing.[8,9] More specific functional behavior has also been demonstrated in metamaterials, including radio frequency (RF) limiting[10] and harmonic generation.[11]

Metamaterials are also well-suited for other functional behaviors, including electromagnetic power harvesting, the focus of this work. Power harvesting devices convert one type of energy to another, typically converting to a direct current (DC) signal. Many types of energy can be harvested, from acoustic (using a piezoelectric harvester)[12] to electromagnetic (using a rectenna).[13] Power harvesting devices require a method to couple to the energy that will be harvested as well as a device to convert the energy from one form to another. By their very nature, metamaterials are designed to couple to various types of energy, e.g., from acoustic[14] to optical,[15] and thus provide a natural platform for power harvesting. Electromagnetic metamaterials provide flexibility in design due to their electrically small, low-profile nature.[16] Since metamaterials are typically designed as infinite arrays, the resonant frequency and input impedance include coupling effects. Metamaterials can be adapted to various applications, such as flexible sheets to cover surfaces.[17] Moreover, many metamaterials that have been presented in the literature require some form of external signal. This could be a DC bias voltage[18] or a large external pump signal.[7] In general, metamaterials could be modified to harvest such an external signal that is already present for other purposes. With these design advantages, power harvesting metamaterials offer design flexibility for a large number of applications that general antenna-based microwave power harvesting devices may lack.

A recent simulation-based study[19] investigated the conversion efficiency between incident RF power and induced power in a split-ring resonator (SRR). Our work is focused on the experimental measurement of RF to DC efficiency based on the conventional effective area of the SRR. We demonstrate that metamaterials can also include embedded devices to convert the incident RF energy to a DC voltage, providing a platform for power harvesting that utilizes the advantages of metamaterial design.

An SRR is a canonical example of a resonant metamaterial particle and is used as the basis for the unit cells of the metamaterial power harvester designed here. By tuning the SRR parameters, we design an SRR (Fig. 1) to resonate at 900 MHz using an S-parameter simulation within Computer Simulation Technology (CST) Microwave Studio software. Using CST Microwave Studio, we can also simulate the effects of embedding devices within the SRR by retrieving its S-parameters using a lumped port. The retrieved S-parameters are loaded into Agilent Advanced Design System (ADS), allowing us to simulate both fullwave 3D effects and circuit-level nonlinear effects.

An SRR couples strongly to an incident magnetic field and can be loaded with a wide variety of circuit elements. In this work, we embed a rectifying circuit within an SRR to convert the incident RF power to DC power. A number of rectifying circuits could be chosen depending on the particular application for the power harvesting metamaterial. We choose to use a Greinacher[20] circuit because the output voltage is double the input voltage maximum, which allows for
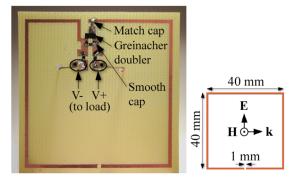
FIG. 1. (Left) Power harvesting SRR. (Right) Plain SRR with dimensions shown, 1 mm traces. Incident wave polarization is also shown.

[a] allen.hawkes@duke.edu

easier power transmission and measurement. A Greinacher circuit has a lower effective capacitance in comparison to other rectification circuits (such as a bridge rectifier), allowing faster switching and thus a higher frequency of operation. A Greinacher circuit also has a low threshold voltage, allowing operation at lower incident power levels. A Greinacher voltage doubler can be placed across a gap (Fig. 1) in the top side of the copper trace to rectify the induced current present in the SRR. Schottky diodes are used for the voltage doubler due to their typically low open junction capacitance and fast switching capabilities, which allow for rectification of a high-frequency RF signal, as well as their typically low threshold voltage. A resistive load placed across the output of the voltage doubler is a simple way to determine DC power out using $P = V^2/R$. This DC power is maximized through ADS for matching capacitor and resistor values. The parameters of our selected Schottky diode (HSMS 2862) and the simulated S-parameters of the above SRR in a model of our actual parallel-plate waveguide are input into ADS for the simulation, shown in a schematic in Fig. 2.
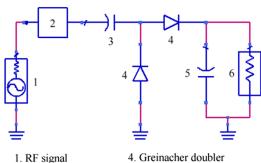
The simulated components lead to a maximum efficiency of 61% for an input power of 24.25 dBM, the maximum available experimental power (at an incident power density of approximately 1.6 mW/cm^2). One important figure of merit for a power harvester is its RF-to-DC power conversion efficiency:

$$\eta = \frac{P_{DC}}{P_{RF}}.$$

We determine P_{RF} by measuring the total incident power in our measurement apparatus. For a large metamaterial sample, we assume that the total incident power to the measurement apparatus is incident on the metamaterial. For a single unit cell, it is necessary to use the effective area of the unit cell to determine the incident P_{RF}. The maximum effective area may be calculated by[21]

$$A_{e,max} = \frac{\lambda^2}{4\pi}D_0,$$

where $D_0 = 1.5$ since the SRR is effectively a small loop illuminated by a transverse electromagnetic (TEM) wave. For an SRR resonant at 900 MHz, the effective area is thus $A_{e,max} = 5.3A_{physical}$. The full waveguide area is

FIG. 3. Placement of SRR within open waveguide.

approximately $6.8A_{physical}$ where $A_{physical}$ is the physical area of a single SRR unit cell. As $A_{e,max}$ is 78% of the open wave guide area, 78% of the incident power density is used as input power for simulation of the single cell.

The designed voltage doubler and resistive load are added to an SRR as shown in Fig. 1, resulting in the power harvesting metamaterial unit cell. To observe power harvesting capabilities, the cell is placed in an open, TEM waveguide (Fig. 3) where input power is produced by a signal generator and amplifier, and output power is measured with an oscilloscope via leads placed across the resistive load (Fig. 4). The DC power harvested is determined by $P = V^2/R$ as previously mentioned, and input power is measured with a spectrum analyzer connected to the signal generator and amplifier via the open waveguide. By increasing the incident power from 13 to 24 dBm and measuring the DC output from the SRR, the normalized harvested power, $P_{DC}/P_{RF,incident}$, as a function of incident power $P_{RF,incident}$ and resulting efficiency are determined at each point. The maximum efficiency of the single cell is 14.2%, setting $P_{RF,incident}$ as 78% of the total input power from effective area calculations.

Multiple power harvesting SRR cells are then tested simultaneously to create the power harvesting metamaterial, which is accomplished through a 5×1 array shown in Fig. 5. Through a parallel connection of the leads from each SRR's resistive load, the total power harvested by the metamaterial is found in the same way as the single cell. The maximum efficiency for the 5×1 array is 36.8%, where P_{in} is the entire input power because the array spans the entire length. Measured efficiencies of both the single and the array of power harvesting SRRs are shown in Fig. 6. Also shown in Fig. 6 is the open circuit voltage, V_{OC}, a load-independent measure of the available voltage harvested by the array of SRRs.

The power harvesting metamaterial array is more efficient than the single unit cell. This is partially due to the

1. RF signal
2. SRR S-params
3. Matching cap
4. Greinacher doubler
5. Smoothing cap
6. Load

FIG. 2. ADS simulation schematic of SRR power harvester. CST Microwave Studio was used to determine the SRR S-parameters.
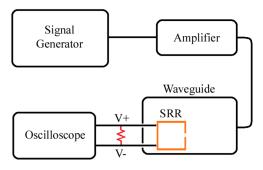
FIG. 4. Experimental test setup schematic.
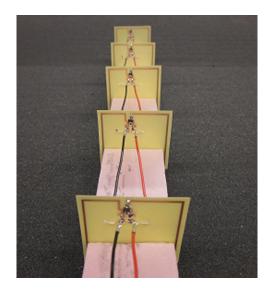
FIG. 5. 5×1 array of power-harvesting SRRs.
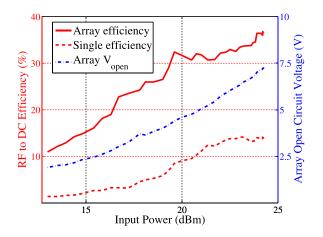
FIG. 6. RF to DC efficiency of power harvesters. The open-circuit voltage of the array is also shown.

FIG. 7. RF to DC efficiency of SRR array as a function of load resistance, for both simulation and experiment. Note that both show maximum efficiency for around 70–80 Ω.

experimental results showing an optimal resistive load for DC power transfer of 70–80 Ω. The SRR power harvester is an example of functional metamaterial that may be suitable for a wide variety of RF applications that require power delivery to any active integrated components.

larger effective area of the array. While the effective area of the single cell is 78% the entire width of the waveguide, some of this power is not harvested by the single unit cell due to fringing effects on the sides of the waveguide. For this reason, placement of multiple cells that together span the entire waveguide width results in a higher efficiency as the array captures more of the electromagnetic energy that undergoes the fringing effect.

Another important relationship is the efficiency of the power harvesting metamaterial as a function of load resistance. Simulated and experimental efficiencies for the 5×1 array are shown in Fig. 7. Though the experimental efficiencies do not match the values from simulation, the simulation does accurately predict the maximum harvested power as falling approximately within the range of 70–80 Ω, confirmed by the experimental data. The experimental efficiency maximum occurs at 70 Ω, and the simulated maximum occurs at 82 Ω, showing close correspondence.

In summary, we have designed, simulated, and experimentally measured a functional metamaterial power harvester capable of converting up to 36.8% of the incident RF power to DC power. Through a parallel connection of five SRRs, a V_{OC} of 7.3 V is achieved. Simulations match

[1] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, Phys. Rev. Lett. **84**, 4184–4187 (2000).
[2] R. W. Ziolkowski, Phys. Rev. E **70**, 046608 (2004).
[3] N. Fang, H. Lee, C. Sun, and X. Zhang, Science **308**(5721), 534–537 (2005).
[4] D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J. B. Pendry, A. F. Starr, and D. R. Smith, Science **314**(5801), 977–980 (2006).
[5] I. V. Shadrivov, S. K. Morrison, and Y. S. Kivshar, Opt. Express **14**(20), 9344–9349 (2006).
[6] H.-T. Chen, J. F. O'Hara, A. K. Azad, A. J. Taylor, R. D. Averitt, D. B. Shrekenhamer, and W. J. Padilla. Nature Photon. **2**(5), 295–298 (2008).
[7] A. R. Katko, S. Gu, J. P. Barrett, B.-I. Popa, G. Shvets, and S. A. Cummer, Phys. Rev. Lett. **105**, 123905 (2010).
[8] D. Huang, A. Rose, E. Poutrina, S. Larouche, and D. R. Smith, Appl. Phys. Lett. **98**(20), 204102 (2011).
[9] A. Rose, D. Huang, and D. R. Smith, Phys. Rev. Lett. **110**, 063901 (2013).
[10] A. R. Katko, A. M. Hawkes, J. P. Barrett, and S. A. Cummer, IEEE Antennas Wireless Propag. Lett. **10**, 1571–1574 (2011).
[11] I. V. Shadrivov, A. B. Kozyrev, D. W. van der Weide, and Y. S. Kivshar, Appl. Phys. Lett. **93**(16), 161903 (2008).
[12] M. Lallart, D. Guyomar, C. Richard, and L. Petit, J. Acoust. Soc. Am. **128**, 2739–2748 (2010).
[13] N. Zhu, R. W. Ziolkowski, and H. Xin, Appl. Phys. Lett. **99**(11), 114101 (2011).
[14] J. Li and C. T. Chan, Phys. Rev. E **70**(5), 055602 (2004).
[15] H. J. Lezec, J. A. Dionne, and H. A. Atwater, Science **316**(5823), 430–432 (2007).
[16] R. W. Ziolkowski and A. Erentok, IEEE Trans. Antennas Propag. **54**(7), 2113–2130 (2006).
[17] G. X. Li, S. M. Chen, W. H. Wong, E. Y. B. Pun, and K. W. Cheah, Opt. Express **20**(1), 397–402 (2012).
[18] J. Zhao, Q. Cheng, J. Chen, M. Q. Qi, W. X. Jiang, and T. J. Cui, New J. Phys. **15**, 043049 (2013).
[19] O. M. Ramahi, T. S. Almoneef, M. Alshareef, and M. S. Boybay, Appl. Phys. Lett. **101**, 173903 (2012).
[20] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 6th edition (Oxford University Press, 2009).
[21] C. A. Balanis, *Antenna Theory: Analysis and Design* (Wiley-Interscience, New York, 2005).

Ambient RF Energy Harvesting in Urban and Semi-Urban Environments

Manuel Piñuela, *Student Member, IEEE*, Paul D. Mitcheson, *Senior Member, IEEE*, and Stepan Lucyszyn, *Senior Member, IEEE*

Abstract—RF harvesting circuits have been demonstrated for more than 50 years, but only a few have been able to harvest energy from freely available ambient (i.e., non-dedicated) RF sources. In this paper, our objectives were to realize harvester operation at typical ambient RF power levels found within urban and semi-urban environments. To explore the potential for ambient RF energy harvesting, a city-wide RF spectral survey was undertaken from outside all of the 270 London Underground stations at street level. Using the results from this survey, four harvesters (comprising antenna, impedance-matching network, rectifier, maximum power point tracking interface, and storage element) were designed to cover four frequency bands from the largest RF contributors (DTV, GSM900, GSM1800, and 3G) within the ultrahigh frequency (0.3–3 GHz) part of the frequency spectrum. Prototypes were designed and fabricated for each band. The overall end-to-end efficiency of the prototypes using realistic input RF power sources is measured; with our first GSM900 prototype giving an efficiency of 40%. Approximately half of the London Underground stations were found to be suitable locations for harvesting ambient RF energy using our four prototypes. Furthermore, multiband array architectures were designed and fabricated to provide a broader freedom of operation. Finally, an output dc power density comparison was made between all the ambient RF energy harvesters, as well as alternative energy harvesting technologies, and for the first time, it is shown that ambient RF harvesting can be competitive with the other technologies.

Index Terms—Ambient RF, energy harvesting, maximum power point tracking (MPPT), multiband, rectenna, RF survey, RF-dc.

I. INTRODUCTION

FOR ALMOST 50 years, far-field RF technology has been used to remotely power systems from relatively large unmanned helicopters [1] to very small smart dust sensors [2] and contact lenses that measure eye pressure [3]. With all these systems, a dedicated RF source is used, where the operator may have control over the effective isotropically radiated power (i.e., both transmit power and antenna characteristics), beam pointing and polarization of the RF source, ensuring optimal line-of-sight operation between the source transmitter (TX) and harvesting receiver (RX). It is important to highlight that this work will

focus only in radiative power transfer and not inductive or near-field power transfer, as demonstrated in [4]. A more convenient solution, however, is to power these devices from ambient RF energy sources, such as television and mobile phone signals, thus removing the need for a dedicated source. As ambient RF levels are lower than those that can be provided by a dedicated RF source, the efficiency of the harvesting system, and its minimum startup power are of critical importance.

In order to assess the feasibility of deploying ambient RF energy harvesters, the available RF power needs to be measured in different locations. Such measurements, in conjunction with knowledge on harvester performance, can then be used to determine the locations at which RF harvester powered devices can be successfully deployed. Several RF spectral surveys, which measure ambient RF power levels from sources such as television and mobile phone base stations, have been previously reported. Many have been undertaken using personal exposimeters or spectrum analyzers, where the exact location of each measurement is unknown and with RF power levels only being reported under general scenarios (e.g., outdoor, indoor, street, bus, etc.) [5], [6]. While being of academic interest for health-related research [7], the lack of power level and specific time/location information limits their usefulness for exploitation in ambient RF energy harvesting applications.

Most rectennas (normally comprising an antenna, impedance matching network, rectifier, storage element, and load) presented in the open literature have been tested using dedicated sources rather than harvesting from ambient RF energy [8]. In recent years, efficiencies as high as 78% [9] and 90% [10] have been demonstrated with relatively high input RF power levels (i.e., $> +10$ dBm). Moderate efficiencies have also been achieved using dedicated TXs that provided relatively low input RF power levels; e.g., an efficiency of 60% was achieved with -22.6-dBm input power [11]. In one demonstrator [12], designed to operate over a broad range of input RF power levels, (-30 to $+30$ dBm), the efficiency increased from 5% at a low input RF power to a peak of 80% at $+25$ dBm.

Despite advancements in end-to-end (i.e., input RF to output dc) power conversion efficiencies at low input RF power levels (similar to those measured in the spectral surveys), only a few attempts at true ambient RF energy harvesting have been reported. For example, one relatively efficient rectenna, utilizing a modified omnidirectional patch antenna, has an efficiency of 18% with a single-tone input RF power of -20 dBm [13]. This dedicated signal source was meant to emulate the input RF power levels measured from a nearby digital TV (DTV) TX in Tokyo, Japan, but did not take into account the more realistic effect of harvesting from a modulated broadband signal. Another attempt

at harvesting ambient RF energy from a mobile phone base station at 845 MHz was reported in [14]. This prototype managed to power an LCD thermometer for 4 min, but only after harvesting for 65 h. In that work, when the authors used a dedicated signal source with a single-tone input RF power of −15 dBm, an efficiency of 3% was recorded. A batteryless location sensor has also been demonstrated [15], powered by a rectenna with a printed antenna on a flexible substrate and a solar cell, although no details for the RF-dc efficiency were reported. Finally, successful prototypes capable of harvesting energy using TV antennas were presented, but again no details of their efficiency were given [16], [17].

In order to demonstrate the feasibility for implementing ambient RF energy harvesting, here we first present the results of a citywide RF spectral survey, indicating suitable locations and associated RF bands with sufficient input RF power density levels for harvesting. Based on these results, rectennas were then fabricated and their efficiencies, under ambient RF energy harvesting operation, were calculated using *in-situ* field strength measurements. Furthermore, an investigation of multiband rectenna arrays is also presented, demonstrating the tradeoffs between series (voltage summing) and parallel (current summing) topologies with the aim of reducing the minimum input power required for harvester operation. Finally, a comparison between measured ambient RF energy harvesting and alternative forms of energy harvesting technologies is presented; highlighting, for the first time, the practical feasibility of exploiting existing freely available sources of RF energy.

II. LONDON RF SURVEY

In order to quantify input RF power density levels present in a typical urban and semi-urban environment, a citywide RF spectral survey within the ultrahigh frequency (0.3–3 GHz) part of the frequency spectrum was conducted within Greater London. A number of citywide RF spectral surveys have previously been conducted, but in general, only a few samples were taken, giving little insight into (semi-)urban environments [14], [18], [19]. Other surveys [20], [21] compare their measurements relative to the distance from the nearest TX. In a (semi-)urban environment, this may not provide enough information about the RF spectrum since there is likely to be local geographical variations in base-station density and propagation characteristics (e.g., multipath effects and diffraction around and attenuation through buildings).

Each station on the London Underground network was used as a survey point to provide a robust dataset for representing Greater London in terms of geographical distribution and population density, having a combination of urban (in the center) and semi-urban (in surrounding areas) characteristics. Measurements were taken at each of the 270 stations (from a randomly chosen exit, at street level and a height of 1.6 m). To provide traceability and for use as a historical reference, time stamps and GPS locations were recorded. In addition, measurements were taken inside a building at Imperial College London (ICL), to represent a typical office block within an urban environment.

A. Methodology

Mobile phone usage varies during the daytime, and hence, ambient RF energy in their bands is expected to be time dependant, with more energy available during the daytime than at night time. Therefore, in order to be able to make fair comparisons between locations, measurements were taken between 10:00 am and 3:00 pm on weekdays over a period of one month (between March 5, and April 4, 2012). Electric field strength was measured between 0.3–2.5 GHz using an Agilent N9912A FieldFox RF analyzer [22] with a calibrated Aaronia BicoLOG 20300 omnidirectional antenna [23]. It is important to note that the spectral measurements were undertaken during the analog-to-digital switchover period in the U.K. and so the measurements for DTV may represent an underestimate of present RF power levels measured now that the switch over is complete [24]. It should also be noted that this survey was conducted prior to the 4G network being switched on within the U.K.

A "panning method," which complies with international regulations for measuring exposure limits, was used [25]–[27]. Here, the calibrated antenna is rotated to three orthogonal axes while the spectrum analyser is set to "max-hold," ensuring that the maximum reading is recorded. For each measurement, more than 1 min was allocated to allow for more than three sweeps across the selected frequency range. Additionally, to maintain a comparable signal-to-noise (S/N) ratio, attenuation was introduced (with a minimum set at 5 dB) to avoid compression when high input RF power levels were detected. For all measurements, the resolution bandwidth (BW) was fixed at 100 kHz, the internal amplifier was turned on and the highest resolution of 1001 points was selected. These settings provide the ability to obtain a snapshot of the power density that can be expected in an urban or semi-urban environment from continuously variable sources.

B. Results

After inputting the manufacturers' frequency-banded antenna factors into the spectrum analyzer, to ensure a fully calibrated system, the electric field strength measurements were taken. The input RF power density (S) is then calculated from the electric field strength measurement. Fig. 1 shows the input RF power density measured outside the Northfields London Underground station, where the spectral bands for DTV, GSM900, GSM1800, 3G, and Wi-Fi can been clearly identified.

A well-designed rectenna should ideally be capable of harvesting energy across an entire band, and thus it is important to calculate the total band power. The banded input RF power density S_{BA} (nW/cm^2) is calculated by summing all the spectral peaks across the band (i.e., in a similar way, the spectrum analyzer calculates channel power). These levels provide a snapshot of source availability at the time and location of the measurement. Moreover, they are used as a harvester design starting point since the power density at each band will define the input impedance of a rectenna.

The exact frequencies for each band are set by the U.K.'s official frequency band allocation [28]; the GSM900, GSM1800 and 3G base transmit (BTx) bands were separated from the associated mobile transmit (MTx) bands. Table I shows average
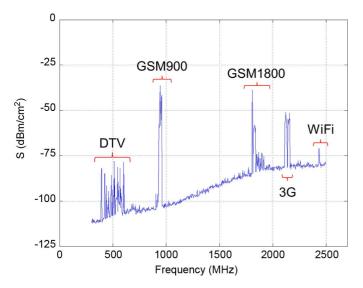
Fig. 1. Input RF power density measurements outside the Northfields London Underground station.

TABLE I
SUMMARY OF LONDON RF SURVEY MEASUREMENTS

Band	Frequencies (MHz)	Average S_{BA} (nW/cm^2)	Maximum S_{BA} (nW/cm^2)
DTV (during switch over)	470–610	0.89	460
GSM900 (MTx)	880–915	0.45	39
GSM900 (BTx)	925–960	36	1,930
GSM1800 (MTx)	1710–1785	0.5	20
GSM1800 (BTx)	1805–1880	84	6,390
3G (MTx)	1920–1980	0.46	66
3G (BTx)	2110–2170	12	240
WiFi	2400–2500	0.18	6

RF power levels across all London Underground stations for the banded input RF power density measurements. It can be seen that all base-station transmit levels are between one and three orders of magnitude greater than the associated MTx levels. For this reason, and the fact that the population of transmitting mobile phones in close proximity of the harvester is highly variable, only base-station TXs will be considered further.

From our London RF survey, DTV, GSM900, GSM1800, 3G, and Wi-Fi were identified as potentially useful ambient RF energy harvesting sources, although DTV appears to be heavily dependent on line-of-sight and sudden changes in atmospheric conditions (e.g., temperature inversion) and Wi-Fi is very dependent on user traffic. It should be noted that the mobile phone base-station TXs employ vertically polarized antennas, placing a constraint on harvester orientation in deployment. With DTV, within the U.K., the main TXs have horizontally polarized antennas, while repeater TXs have vertically polarized antennas.

It is convenient to define the boundary between urban and semi-urban environments by the line that separates zones 3 and 4 on the London Underground map [29]. As one would expect, the central zones 1–3 host the highest density of base stations. As shown in Table II, a banded input RF power density threshold was selected to filter the ten London Underground stations with the highest measurements for each band. With DTV, the highest

TABLE II
INPUT RF POWER DENSITY THRESHOLD: LONDON UNDERGROUND STATIONS WITHIN CENTRAL ZONES 1–3 (URBAN) AND OUTER ZONES 4–9 (SEMI-URBAN)

Band	S_{BA} Threshold (nW/cm^2)	Number of Stations Urban	Number of Stations Semi-urban
DTV (during switch over)	40	10	0
GSM900	230	8	2
GSM1800	450	7	3
3G	62	6	4
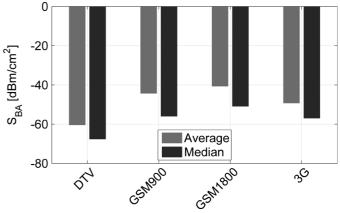
Fig. 2. Banded input RF power density measurements for the four largest ambient sources in Greater London.

recorded measurements were all found within the urban environment. This is because Greater London's main DTV TX (at Crystal Palace) is located on the southeastern boarder of zones 3 and 4 and there are no London Underground stations further south. With mobile phones, more than 50% of the stations were inside the urban environment and those in a semi-urban environment were all located in close proximity to a cluster of base-station TXs.

Using the complete dataset from the London RF survey, Fig. 2 shows the average and median of the banded input RF power density measurements for the four largest ambient RF sources in Greater London. It can be seen that more than half of the locations have below average power levels. This is due to the fact that several stations had maximum levels that were considerably higher than the average because of their close proximity to TV TXs (e.g., Crystal Palace), extremes in base-station density and propagation characteristics.

In addition to the London RF survey, measurements within the Department of Electrical and Electronic Engineering building at ICL were taken on the 11th floor of the south stairwell. These are shown in Table III. As can be seen, DTV and GSM900 have a higher than average power level, due to a near line of sight from the TV TX and a close proximity to the 2G GSM900/1800 base stations.

The dataset from the London RF survey, with all relevant information (e.g., locations, timestamps, and banded input RF power density measurements), can be found at our interactive website: www.londonrfsurvey.org [30]. These measurements were used to design efficient harvesters and compared to ICL

TABLE III
MEASURED BANDED INPUT RF POWER DENSITIES AT ICL

	DTV	GSM900 (BTx)	GSM1800 (BTx)	3G (BTx)
S_{BA} [nW/cm^2]	18	48	50	3

measurements to identify locations in Greater London where the designed harvesters could operate. The design procedures and prototype test measurements will be presented in the following sections.

III. SINGLE-BAND AMBIENT RF ENERGY HARVESTERS

In order to implement efficient ambient RF energy harvesters, designed for the banded input RF power density levels measured at ICL, a set of single-band prototypes were realized and characterized; these will be compared to multiband array architectures in Section V.
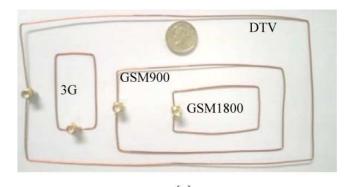
A. Antenna Design and Measurements

Since our harvesters are intended to operate within a general (semi-)urban environment, where the exact location of the TX source is unknown, the rectennas' antennas need to be as close to omnidirectional as possible, avoiding the need for beam-pointing during deployment. This is at the obvious expense of limited antenna gain, and therefore, the corresponding levels of P_{RF} that the rectifier can receive. Conversely, if the location of the TX is known, then it may be tempting to use a high gain antenna, but this would require an appropriate level of beam-pointing and polarization matching that can be established and maintained.

Another requirement is that the antennas need to be easily scalable across all frequency bands since one important objective for this work is to compare and contrast different banded harvesters. Finally, the antennas need to be easily fabricated. For all these reasons, a linear polarized folded dipole was selected, although a monopole would also be suitable [31].

To simplify impedance matching between the antenna and rectifier, a modified folded dipole was used to obtain the required 50-Ω reference input impedance. A balun does not need to be employed, as there is no significant degradation in performance for this particular application, even with the use of an unbalanced microstrip rectifying circuit [32]. Furthermore, the antenna was not integrated onto a substrate, to give the additional freedom to embed the harvester on windows or within walls, furnishings, fixtures, or fittings. To this end, two different antennas were fabricated for each band; one made using a 560-μm diameter copper wire and the other with 75-μm-thick 25-mm-wide copper tape. The fabricated antennas are shown in Fig. 3. Since the copper tape was not rigid enough to retain its shape, it was placed on a Perspex substrate, to represent a flat panel.

To design the antennas, full-wave 3-D electromagnetic simulations were performed using CST Microwave Studio. As discussed previously, the antennas were designed to be as omnidirectional as possible, while covering as much of the ambient RF source BW as possible. Fig. 4 shows the typical simulated gain profile for the DTV tape antenna, having a front-to-back ratio

(a)

(b)
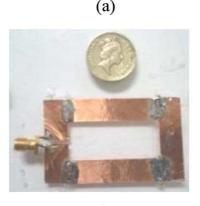
Fig. 3. 50-Ω folded-dipole antennas shown next to a British £1 coin. (a) DTV, GSM900 (BTx), GSM 1800 (BTx) and 3G (BTx) copper wire antennas. (b) 3G (BTx) copper tape antenna on Perspex.
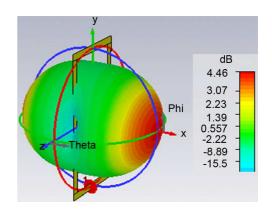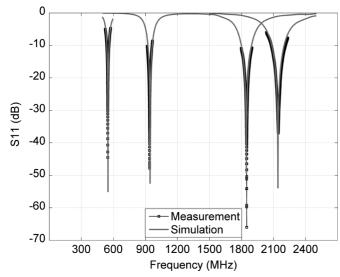
Fig. 4. Simulated beam profile for the DTV tape antenna.

of unity. Table IV shows the simulated gain and the 10-dB return-loss fractional BW for the optimized copper wire and tape antennas.

Fig. 5 shows excellent agreement between predicted and measured return-loss results, within a 10-dB return-loss bandwidth, for the eight fabricated single-band antennas. As one would expect with such a simple classical antenna, the out-of-band performances (not shown) were also in good agreement. It was found that better return-loss measurements are achieved with our single-band antennas when compared to other reported single-band omnidirectional [13] and multiband [33] designs. The latter may be important, as it may be tempting to implement a more compact multiband rectenna, but which

TABLE IV
SIMULATED GAIN AND 10-dB RETURN LOSS FRACTIONAL BW FOR
FOLDED-DIPOLE SINGLE-BAND ANTENNAS

Band		Wire		Tape	
	BW (%)	Gain (dBi)	BW (%)	Gain (dBi)	BW (%)
DTV	26	4.35	4	4.48	6
GSM900	3.7	4.42	4.3	4.73	4.3
GSM1800	4.1	4.32	5.3	4.73	10.7
3G	2.8	4.39	5.4	4.76	12

Fig. 6. Predicted input RF power levels for the four largest ambient sources at the ICL testing location.

only covers approximately 35% of the target frequency range. In other work [13], 5-dB return-loss fractional bandwidth is adopted for RF energy harvesting applications.

The fractional bandwidth of the antennas having a minimum return loss of 5 dB is too great to assume a constant antenna gain over the whole band. Therefore, an additional advantage of using 10-dB return-loss fraction bandwidths is that (1) can be used to calculate the input RF power with the assumption that the midband antenna gain is constant with frequency [34]. Therefore, the time-averaged input RF power $P_{\rm RF}$ is given by

$$P_{\rm RF} = S_{BA} \cdot A_{\rm real} \text{ and } A_{\rm real} \approx G(f_o)\frac{\lambda_o^2}{4\pi} \qquad (1)$$

where $A_{\rm real}$ is the real aperture (or capture area) of the antenna, λ_o is the free-space wavelength at the midband frequency f_o, and $G(f_o)$ is the rectenna's antenna gain at f_o.

Substituting the measured banded input RF power densities recorded in Table III and the predicted midband antenna gains in Table IV into (1), realistic values for $P_{\rm RF}$ can be calculated for all four bands, with the results shown in Fig. 6. It can be seen that with all antenna gains being in the region of ~ 4.5 dBi, the 2G GSM900/1800 harvesters will generate the highest input RF power levels, due to the high-banded input RF power density levels measured *in situ*. At the other extreme, the 3G harvesters will be the worst performers. As only a small fraction of the required frequency range is covered by the DTV antennas, the predicted values for $P_{\rm RF}$ represent an overestimation.

B. Rectifier Design and Measurements

Based on a previously reported analysis [35] and the predicted input RF power levels presented in Fig. 6, the zero-bias SMS7630 diode (in a series configuration) was selected as the optimal solution for our ambient RF energy harvesters, as shown in Fig. 7. In a series configuration, the junction capacitance $(C_j(V))$ of the diode dominates the detector's impedance, as long as $C_{\rm out} > C_j(V)$, and thus $C_{\rm out}$ has little or no effect on the matching circuit. This allows $C_{\rm out}$ to be large enough to provide a ripple-free output voltage. In contrast, $C_{\rm out}$ must be less than 1 pF to achieve a good impedance match with a shunt configuration as $C_{\rm out}$ appears in parallel with $C_j(V)$ and the packaging parasitic capacitance. However, $C_{\rm out}$ is too small to

Fig. 5. Input return-loss predictions and measurements for all single-band folded-dipole antennas. (a) Wire. (b) Tape.

may ultimately not give better ambient RF energy harvesting performance.

Obtaining a minimum acceptable return loss over an antenna fractional bandwidth as large as that of the source is key to harvest as much input RF energy as possible. As can be seen in Table IV, where the fractional bandwidth is defined for a 10-dB return loss, our antennas have a fractional bandwidth greater than those of the sources, with the exception of DTV, which

Fig. 7.   Series detector configuration with L-matching network.

Fig. 8.   Input return-loss predictions and measurements for all first prototype single-band rectifiers, with $P_{RF} = -20$ dBm at the input and optimal load resistances at the output.

provide a ripple-free dc voltage to the load. This can be overcome with a matching network that will allow a good impedance match with a large output capacitor, but at the expense of introducing losses. Furthermore, as shown in [36], these issues start to become negligible with a shunt configuration as the shunt diode becomes more self-biased as the input power increases. Simulations were performed using Agilent Technologies' ADS software. Its Momentum package not only takes into account the losses from the low-cost FR4 substrate, but also calculates fringing fields, effects from which are passed on to the harmonic-balance package for simulating the nonlinear behavior of the rectifier.

A good impedance match was achieved by employing a simple matching network; a series lumped-element inductor was used to absorb part of the capacitive reactance from the series diode and an additional quarter-wavelength short-circuit shunt stub was employed to achieve the desired 50-Ω impedance [37]. Since the impedance of the diode varies with frequency and input RF power, impedance matching between the antenna and rectifier was first undertaken by finding the optimal output load resistance for an input RF power level of -20 dBm with a single-tone source at the midband frequency. After the optimal load was found, further broadband optimization was performed to the matching network and the load to ensure good impedance matching throughout the target frequency range and the measured P_{RF} for each band.

As with the antenna analysis, and unlike conventional RF circuits that adopt the more traditional half-power bandwidth definition, the rectifier should adopt the 10-dB-input return-loss bandwidth. The reason for this is that, for ambient RF energy harvesting applications, the input RF power is at a premium and so what little energy is available should not be wasted by being reflected back from avoidable impedance mismatches at either the antenna or rectifier.

Fig. 8 shows excellent agreement between predicted and measured input return loss results, within the -10-dB bandwidth, for the DTV and GSM900 rectifiers, having fractional bandwidths of 5.7% (below target) and 4.8%, respectively. With these lower frequency designs, the fundamental and higher order harmonics were below -55 dBm, ensuring a clean dc voltage at the load, without the need for any output filtering. Reasonable agreement was found with the GSM1800 and 3G rectifiers, having fractional bandwidths of 1.6% (below target) and 7.4%, respectively. It was found that with these two higher frequency designs, the higher order harmonics were -40 dBm at the output. This reduced performance, as illustrated in Fig. 8, is due to the higher series inductive reactance leads of the output shunt storage capacitor. For this reason, a
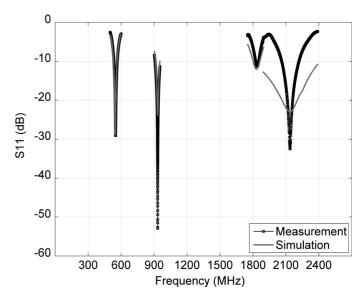
second prototype version (v2) was designed for the 3G rectifier, using distributed-element components for the input impedance matching stage and an additional output filter stage. With the lumped-element matching network, it was found that in order to achieve good input impedance matching to 50 Ω, all the microstrip transmission lines had to have a characteristic impedance of 92 Ω. With the distributed-element matching network, a simple shunt quarter-wavelength open-circuit stub, designed for operation at the fundamental frequency, was employed. A microstrip line was added between the cathode of the diode and the stub to absorb the capacitive reactance of the diode. The stub effectively filters to < -50 dBm the higher order harmonics. The microstrip design can be seen in Fig. 13.

C. PMM

Since the input RF power from ambient sources can be represented as a multi-tone source, with power levels fluctuating across the target frequency range, the output impedance of the rectifier is time varying. A power management module (PMM) capable of performing maximum power point tracking (MPPT) is required.

For our work, a low-power integrated-circuit PMM from Texas Instruments Incorporated (BQ25504) was selected, due to its low quiescent current (< 330 nA) and low input voltage operation (~ 80 mV hot-start and 330-mV cold-start) [38]. It is worth noting that its startup voltage is lower than PMMs previously reported and realized using hybrid circuits for RF energy harvesting [13]. The BQ25504 PMM includes a boost converter that steps up its input voltage (having a 350-mV average value during ambient operation) to useful levels between 2.4–5.3 V. The BQ25504 also has a built-in battery management module, which is used to control the duty cycle of the output power to the load.

MPPT operation on the BQ25504 is achieved by periodically sampling the open-circuit voltage (OCV) at the input of the converter, which then draws a current causing the converter

Fig. 9. System block diagram.

Fig. 10. End-to-end efficiencies for ambient RF energy harvesting at ICL.
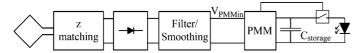
input voltage to fall and be held at a pre-programmed fraction of the OCV (set by a potential divider). In a simple dc circuit, with a resistive source impedance, the optimal ratio is 0.5. For the rectenna-based system, a ratio of 0.48–0.53 was found to maximize the power output of the system. The BQ25504 is designed to charge a storage element, and in this case, a capacitor C_{storage} was used. The programmed PMM continuously charges the storage capacitor, and the load (a low-power LED) was automatically connected to the storage capacitor when the capacitor voltage reaches an upper limit $V_{\text{high}} = 2.84$ V and automatically disconnected when it reaches a lower limit $V_{\text{low}} = 2.40$ V. The duty cycle of the LED can then be used to calculate the efficiency of the system, as will now be described. A diagram of the system is shown in Fig. 9.

IV. END-TO-END EFFICIENCY ANALYSIS

The efficiency η of an RF energy harvesting system is

$$\eta = \frac{P_{\text{dc}}}{P_{\text{RF}}} \qquad (2)$$

where P_{dc} is the time-averaged output (i.e., equivalent dc) power into the storage element (e.g., battery or supercapacitor) and load and P_{RF} is as previously defined. Measurements for this type of system are usually performed in a controlled environment (e.g., an anechoic chamber or TEM cell), using a dedicated constant or variable amplitude single-tone RF signal source [32], [39]. However, the former is not suitable for evaluating ambient RF energy harvesting operation, which has a much broader spectrum of nonconstant input frequencies and where the instantaneous input RF power is time variant. The use of a constant single-tone dedicated source provides a convenient stable reference power to the harvester; while the latter reflects a more realistic signal source having fluctuating power levels across a nonzero bandwidth, multipath, and reflection effects which are very difficult to emulate in a controlled environment.

Therefore, to determine the overall end-to-end efficiency $\eta_{\text{e-e}}$ for a complete ambient RF energy harvester, the input RF energy U_{RF} was calculated based on the harvester's antenna characteristics and the actual banded input RF power density measurements taken at the time of harvester operation, using the Agilent Fieldfox and the calibrated antenna. It is important to note that since the impedance mismatch between the antenna and detector is not taken into account, U_{RF} is higher than expected, providing an underestimate of end-to-end efficiency. The output dc energy U_{dc} was then calculated by measuring the charge–discharge cycle time, t_{cycle} of the storage capacitor between V_{high} and V_{low}, as the LED is repeatedly connected and disconnected. The output dc energy equation is already taking into account the efficiency of the PMM given the fact that the measurements
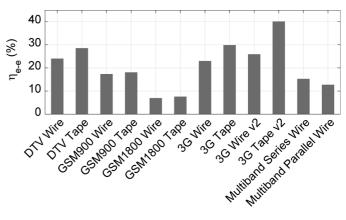
are taken at its output voltage. The end-to-end efficiency of one charge–discharge cycle of C_{storage} is

$$\eta_{\text{e-e}} = \frac{U_{\text{dc}}}{U_{\text{RF}}} \qquad (3)$$

where the input RF energy is given by integrating the time-averaged input RF power over a cycle time, as

$$U_{\text{RF}} = \int_0^{t_{\text{cycle}}} P_{\text{RF}} dt \qquad (4)$$

and the output dc energy is given by the energy supplied to the load, as follows:

$$U_{\text{dc}} = C_{\text{storage}} \frac{(V_{\text{high}}^2 - V_{\text{low}}^2)}{2}. \qquad (5)$$

A. ICL Field Trials

Four single-band ambient RF energy harvesters were assembled by connecting the rectifiers to the wire/tape antennas and PMMs programmed for the optimal load. A 100-μF shunt capacitor was employed as the storage element, providing $U_{\text{dc}} = 115$ μJ. Our system is capable of cold-starting the boost converter and MPPT since the rectenna is capable of providing the minimum starting voltage of 330 mV. When the minimum voltage is reached, the boost converter and MPPT start to operate and the charge–discharge cycle at the load begins, causing the LED to flash. During field trials, t_{cycle} took up to 170 s for the harvester with lowest banded input RF power density, corresponding to 3G with the wire antenna. Table V summarizes the results where t_c and t_d are the charge and discharge times, respectively, and ΣV and ΣI are the multiband voltage and current summing array architectures, respectively. A detailed discussion on the multiband rectenna arrays will be presented in the following section. The end-to-end efficiency was calculated using (3) with data from Fig. 6 and measuring the charge–discharge cycle time during harvesting operation.

Fig. 10 shows the overall end-to-end efficiencies for all the harvester demonstrators, deployed and tested at ICL. As predicted by simulations, the improved 3G v2 demonstrator with

TABLE V
HARVESTERS CHARGE AND DISCHARGE TIMES (t_c, t_d, RESPECTIVELY) FOR A SPECIFIED LOAD

Band	Wire					Tape				
	t_c (s) load independent	t_d (s) load dependant	t_{cycle} (s) load dependant	$P_{dc}(t_d)$ (µW)	$P_{dc}(t_{cycle})$ (µW)	t_c (s) load independent	t_d (s) load dependant	t_{cycle} (s) load dependant	$P_{dc}(t_d)$ (µW)	$P_{dc}(t_{cycle})$ (µW)
DTV	26	12	38	9.6	3	14	18	32	8.2	3.6
GSM900	14	10	24	11.5	4.8	8	13	21	14.4	5.5
GSM1800	43	15	58	7.7	2	22	27	49	5.2	2.4
3G v2	167	3	170	38.4	0.7	96	5	101	1.2	1.1
Multiband ΣV	43	7	50	66	2.3	-	-	-	-	-
Multiband ΣI	55	5	60	92.2	2	-	-	-	-	-

tape antenna outperformed its original design by 11%; achieving an end-to-end efficiency of 40% with an input RF power of only −25.4 dBm.

It is believed that a much greater efficiency can be achieved for the DTV harvester if the fractional bandwidths for the first prototype circuits (i.e., 4.4/4.5% for the antennas and 5.8% for the rectifier) could be increased to match the much greater target value of 26%. Likewise, the reduced efficiency of the GSM1800 harvesters can be attributed to the detrimental effects of the narrowband input impedance matching of the rectifier (i.e., having a fractional bandwidth of only 1.6%, when compared to its target value of 4.1%). Finally, with all the harvesters, the end-to-end efficiencies can be enhanced through better antenna design and optimal polarization matching.

Table VI, shows the number of locations from the London RF survey that would be able to support our harvesters. Unlike the single-band 3G harvester, which can operate at 45% of the locations, our DTV harvester can only be used at two locations (one in zone 2 and the other in zone 3). Therefore, for the general deployment of an ambient RF energy harvester within an (semi-)urban environment, at street level, the single-band DTV harvester may not be practical

V. ARRAY ARCHITECTURES

Since ambient input RF power levels can be low (i.e., below −25 dBm) and dependent on both time and spatial considerations, harvesters could be designed to extract energy with spatial-diversity within the same frequency band or using different frequency bands. For example, with the former, at a particular location there may be only one band that has significant levels of RF energy worth harvesting. In this case, spatial-diversity array architectures may provide more usable output power. Alternatively, with the latter, multiband array architectures may provide more robust operation.

With both forms of parallel array architecture (i.e., spatial-diversity and multiband), a further classification can be seen through the use of either diversity/band switching or a summing node. With the former, physical switches automatically select whichever signal path delivers the highest input RF power level;

TABLE VI
NUMBER OF LOCATIONS FROM THE LONDON RF SURVEY CAPABLE OF
SUPPORTING IDENTICAL HARVESTERS AT THE SAME EFFICIENCY LEVELS

	DTV	GSM900 (BTx)	GSM1800 (BTx)	3G (BTx)
Stations with higher S_{BA}	2	28	68	122

with the latter, power from all signals is combined. Fig. 11 illustrates generic forms of parallel array architectures, showing that switching/summing can be performed electromagnetically at a single antenna or at the output from multiple antennas, rectifiers, or PMMs.

Multiband array architectures, similar to those in Fig. 11(c) and (d), capable of RF harvesting from the four previously identified bands, were selected as possible optimal solutions, given no size/cost constraints. Our objectives were to reach the minimum cold-start voltage at the lowest possible input RF power levels and increase the harvesters' operational capabilities within (semi-)urban environments.

To this end, two different multiple rectenna architectures were investigated. The first with a single shared PMM and the second with multiple PMMs, as illustrated in Fig. 12. To simplify assembly, the wire antennas were selected since they did not require a substrate. Unwanted coupling between the single-band antennas was minimized by placing them a minimum distance of $\lambda_L/5$ apart; where λ_L is the wavelength of the lowest frequency band antenna [40]. For example, the DTV and the 3G antennas were kept at least 11 cm apart, as shown in Fig. 13. This allowed S_{11} measurements to be the same as in Fig. 5 once all antennas were assembled into the array.

A. Multiple Rectennas With a Shared PMM

In order to improve the cold-start performance of the system, the outputs of multiple rectennas can be connected in series, as shown in Fig. 12(a). This increases the probability of the voltage on the input of the PMM reaching the cold-start level (330 mV for the BQ25504) under any given scenario. While cold-starting the PMM, each rectenna harvests (albeit not optimally). Once

(a)
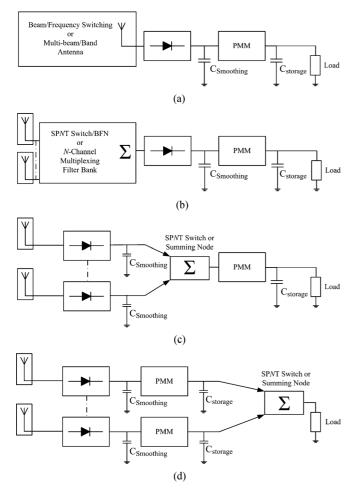
Fig. 11. Parallel array architectures with switching/summing at the: (a) antenna, (b) output of multiple antennas, (c) output of multiple rectifiers, and (d) output of multiple PMMs.

Fig. 12. Multiband array architectures (with $N = 4$ bands). (a) Voltage summing at the outputs of the single-band rectennas. (b) Current summing at the outputs of the single-band harvesters.
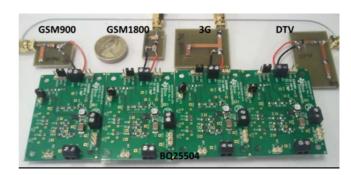
Fig. 13. Rectenna array architecture with individual PMMs for the four largest contributors with wire antennas.

the PMM circuit starts, with the MPPT operating, the harvested power level increases.

The behavior of the series rectenna topology with a shared PMM requires some discussion. As the output impedance and the OCV for each rectenna is different, since they operate at different frequencies and input RF power levels, the rectennas are forced to share the same output current in a series configuration, which does not allow them all to operate at their individual maximum power points. This causes the voltage on each rectenna output, except the one having the highest input RF power, to collapse. This operation is analogous to the partial shading problem with a series string of solar panels [41] sharing a common boost converter. With this photovoltaic system, bypass diodes placed around individual cells stop the poorly lit cells contributing a negative voltage (and power) to the string. In our case, the series circuit formed by the loop antennas and rectifying diodes performs the same task. This means that while all rectennas contribute to system startup, only the rectenna with the highest input RF power contributes significant power for continuous operation once the PMM starts. Fig. 10 shows the end-to-end efficiency for the voltage summing multiband harvester array when tested at ICL. An efficiency of only 15% was achieved with a combined input RF power of -12 dBm. The lower efficiency, when compared to a single-band harvester, is due to the imbalance of rectifier outputs, as discussed above. Here, the charge time was 43 s, compared to 167 s with the lowest contributing single-band 3G harvester with wire antenna.

B. Multiple Rectennas With Individual PMMs

In order to overcome the balancing issues when multiple rectennas share a common PMM, as discussed previously, each rectenna can have its own PMM, whose outputs can be connected to a common storage element (in this case, a 400-μF shunt capacitor, providing $U_{dc} = 461\ \mu$J), as illustrated in Fig. 12(b) and shown in Fig. 13. Although not achieving cold-start as quickly as the series topology, this parallel topology has the advantage of being able to run each rectenna at its maximum power point. In addition, once one rectenna is able to harvest enough energy for a cold-start, all PMMs will start because they share a common storage element, allowing the rectennas with low-input RF power levels to harvest at levels below which they could not do so independently.

This parallel topology was tested and found to be capable of operating in many locations where the series array was unable to operate; e.g., if only one of the bands had $P_{in} > -25$ dBm. As expected, the largest contributor hot-started the other PMMs, allowing them to harvest at an input RF power level down to -29 dBm.

However, as with the previous results for voltage summing, having a combined input RF power of -12 dBm, the efficiency using multiple PMMs is slightly lower, at 13%, as shown in Fig. 10. This is because useful dc output power from the cold-
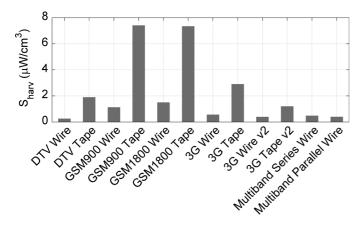
Fig. 14. Output dc power density for all harvesters at ICL.

Fig. 15. Output dc power density comparison for alternative ambient harvesting technologies [40]–[43] against the best current generation of RF harvesters at ICL.

starting harvester is being supplied to the other harvesters for hot-starting, even though some of them may not actually be contributing any of their own harvested power.

VI. OUTPUT DC POWER DENSITY COMPARISON

The volumetric output dc power density S_{harv} (μW/cm^3) represents an important figure of merit for comparing alternative energy harvesting technologies. For ambient RF energy harvesting, the output dc power is calculated by multiplying the effective input RF power by the overall end-to-end-efficiency. The total volume (including that of the antenna, rectifier, and PMM; not including energy storage, as this does not directly affect the dc power output) must be determined. It is important to note that the volume for the antenna could effectively disappear if it is assembled onto a window or within a wall, furnishing, fixture, or fitting. Moreover, the required PMM printed circuit board (PCB) size used throughout these calculations was assumed to be ten times the size of the BQ25504 chip, to account for any necessary additional components.

Fig. 14 shows the output dc power density for all the harvesters demonstrated here. It can be seen that the 2G GSM900/1800 harvesters with tape antennas both have the highest value of $S_{\mathrm{harv}} = 7.4$ (μW/cm^3), when tested at ICL, due to the high-banded input RF power density S_{BA}. The value for the most efficient harvester (i.e., 3G v2 with tape antenna) was not the highest in terms of output RF power because S_{BA} in this band was more than an order of magnitude lower than with GSM900.

S_{harv} allows a direct and meaningful comparison to be made with other alternative energy harvesting technologies. Our best performing ambient RF energy harvester (i.e., GSM900 with tape antenna) was compared against alternative energy harvesting technologies, assuming they used the same PMM board size [42]–[44].

It can be seen in Fig. 15 that ambient RF energy harvesting has a low output dc power density when compared to alternative energy harvesting technologies, but only when the total volume of the first prototype demonstrator is considered. However, when the antenna is absorbed onto or into a background
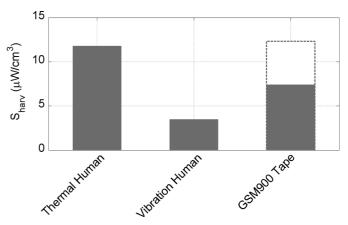
feature and when the PMM is fully integrated into the rectifier, it can outperform (as indicated by the dotted column) the alternative energy harvesting technologies, while providing a complimentary means of extracting energy from the environment. The RF harvesters, however, have the additional advantage in that they do not require a thermal gradient, and unlike vibration-driven devices, they have no moving parts.

VII. CONCLUSIONS

Our objectives were to reach the lowest possible ambient input RF power levels and extend the harvesters' operational capabilities within (semi-)urban environments. To this end, a comprehensive citywide RF spectral survey was undertaken, indicating that more than 50% of the 270 London Underground stations are suitable locations for the deployment of our ambient RF energy harvesters. It has been demonstrated that single-band harvesters can operate with efficiencies of up to 40% in a (semi-)urban environment, and can start to operate from power levels as low as -25 dBm.

To increase the freedom of operation, multiband array architectures were investigated. With the current summing harvester arrays, RF harvesting was achieved at an input RF power level as low as -29 dBm, without any external dc power supply to hot-start the PMM. Limitations on the multiband array architectures were discussed, highlighting the need for further work in balancing rectennas with voltage summing rectenna arrays when operating at lower input RF power levels.

Finally an output dc power density comparison against alternative energy harvesting technologies has shown that RF harvesting can represent a competitive solution within (semi-)urban environments, especially when the antenna can be absorbed into background features.

REFERENCES

[1] "Demonstration of microwave power transmission," *Pract. Wireless*, pp. 835–835, 1965.

[2] T. Salter, G. Metze, and N. Goldsman, "Parasitic aware optimization of an RF power scavenging circuit with applications to smartdust sensor networks," in *IEEE Radio Wireless Symp.*, 2009, pp. 332–335.

[3] "SENSIMED Triggerfish®—Continuous IOP monitoring," SENSIMED, Lausanne, Switzerland, Mar. 2012. [Online]. Available: http://www.sensimed.com/S-Trig-glaucoma.htm, , "

[4] M. Pinuela, D. C. Yates, S. Lucyszyn, and P. D. Mitcheson, "Maximizing DC-to-load efficiency for inductive power transfer," *IEEE Trans. Power Electron.*, vol. 28, no. 5, pp. 2437–2447, May 2013.

[5] W. Joseph, G. Vermeeren, L. Verloock, M. M. Heredia, and L. Martens, "Characterization of personal RF electromagnetic field exposure and actual absorption for the general public," *Health Phys.*, vol. 95, no. 3, pp. 317–330, Sep. 2008.

[6] G. Thuroczy, F. Molnar, J. Szabo, G. Janossy, N. Nagy, G. Kubinyi, and J. Bakos, "Public exposure to RF from installed sources: Site measurements and personal exposimetry," in *1st Eur. Antennas Propag. Conf.*, 2006, pp. 1–4.

[7] P. Elliott, M. B. Toledano, J. Bennett, L. Beale, K. de Hoogh, N. Best, and D. J. Briggs, "Mobile phone base stations and early childhood cancers: Case-control study," *Br. Med. J.*, vol. 340, no. 1, pp. 1–7, Jun. 2010.

[8] N. Shinohara, "Power without wires," *IEEE Microw. Mag.*, vol. 12, no. 7, pp. S64–S73, Dec. 2011.

[9] J. Y. Park, S. M. Han, and T. Itoh, "A rectenna design with harmonic-rejecting circular-sector antenna," *IEEE Antennas Wireless Propag. Lett.*, vol. 3, pp. 52–54, 2004.

[10] "Dengyo rectennas," Nihon Dengyo Kosaku Company Ltd., Tokyo, Japan, Jul. 2012. [Online]. Available: http://www.den-gyo.com/solution/solution01.html

[11] T. Le, K. Mayaram, and T. Fiez, "Efficient far-field radio frequency energy harvesting for passively powered sensor networks," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1287–1302, May 2008.

[12] V. Marian, B. Allard, C. Vollaire, and J. Verdier, "Strategy for microwave energy harvesting from ambient field or a feeding source," *IEEE Trans. Power Electron.*, vol. 27, no. 11, pp. 4481–4491, Nov. 2012.

[13] C. Mikeka, H. Arai, A. Georgiadis, and A. Collado, "Dtv band micropower RF energy-harvesting circuit architecture and performance analysis," in *IEEE Int. RFID-Technol. Appl. Conf.*, 2011, pp. 561–567.

[14] S. Kitazawa, H. Ban, and K. Kobayashi, "Energy harvesting from ambient RF sources," in *IEEE MTT-S Int. Microw. Symp. Workshop*, 2012, pp. 39–42, Series on Innovative Wireless Power Transmission: Technol., Syst., Appl..

[15] R. Vyas, V. Lakafosis, and M. Tentzeris, "Wireless remote localization system utilizing ambient RF/solar power scavenging RFID tags," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 23–28, 2010, pp. 1764–1767.

[16] H. Nishimoto, Y. Kawahara, and T. Asami, "Prototype implementation of ambient RF energy harvesting wireless sensor networks," *IEEE Sens. J.*, pp. 1282–1287, Nov. 2010.

[17] A. Sample and J. R. Smith, "Experimental results with two wireless power transfer systems," in *IEEE Radio Wireless Symp.*, 2009, pp. 16–18.

[18] V. Nitu, G. Lojewski, and S. Nitu, "Electromagnetic field evaluation on an antennas shared site, EUROCON 2009," in *IEEE EUROCON'09*, 2009, pp. 70–75.

[19] Y. Kawahara, K. Tsukada, and T. Asami, "Feasibility and potential application of power scavenging from environmental RF signals," in *IEEE AP-S Int. Symp.*, 2009, pp. 1–4.

[20] H. J. Visser, A. C. F. Reniers, and J. A. C. Theeuwes, "Ambient RF energy scavenging: GSM and WLAN power density measurements," in *38th Eur. Microw. Conf.*, 2008, pp. 721–724.

[21] T. G. Cooper, S. M. Mann, M. Khalid, and R. P. Blackwell, "Public exposure to radio waves near GSM microcell and picocell base stations," *J. Radiol. Protection*, vol. 26, no. 2, pp. 199–211, 2006.

[22] "FieldFox RF Analyser N9912A 4/6 GHz," Agilent Technol., Santa Clara, CA, USA, Tech. Overview, 2012.

[23] "Biconical EMC broadband antennas—BicoLOG Series," Aaronia, Strickscheid, Germany, Data Sheet, 2012.

[24] "Digital TV switchover in the UK," Digital UK Ltd., London, U.K., 2012.

[25] "International commission on non-ionizing radiation protection, guidelines for limiting exposure to time- varying electric, magnetic, and electromagnetic fields (up to 300 GHz)," *Health Phys.*, vol. 74, no. 4, pp. 494–522, 1998.

[26] M. Riederer, "EMF exposure due to GSM base stations: measurements and limits," in *IEEE Int. Electromagn. Compat. Symp.*, 2003, vol. 1, pp. 402–405.

[27] "Revised ECC recommendation (02)04: Measuring non-ionising electromagnetic radiation (9 kHz–300 GHz)," ECC within the Eur. CEPT, Oct. 2003.

[28] Nat. Freq. Planning Group, "United Kingdom frequency allocation table, issue no. 16," 2010, Cabinet Official Committee on U.K. Spectrum Strategy.

[29] Transport for London, "London underground map," Aug. 2012. [Online]. Available: http://www.tfl.gov.uk/gettingaround/1106.aspx

[30] Imperial College London, "London RF Survey," Aug. 2012. [Online]. Available: http://londonrfsurvey.org/

[31] R. Vyas, V. Lakafosis, M. Tentzeris, H. Nishimoto, and Y. Kawahara, "A battery-less, wireless mote for scavenging wireless power at UHF (470–570 MHz) frequencies," in *IEEE AP-S Int. Symp./USNC/URSI Nat. Radio Sci. Meeting*, Jul. 3–8, 2011, pp. 1069–72.

[32] T. Ungan and L. M. Reindl, "Harvesting low ambient RF-sources for autonomous measurement systems," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, 2008, pp. 62–65.

[33] A. Costanzo, A. Romani, D. Masotti, N. Arbizzani, and V. Rizzoli, "RF/baseband co-design of switching receivers for multiband microwave energy harvesting," *Sensors Actuators A, Phys.*, vol. 179, pp. 158–168, Jun. 2012.

[34] J. L. Volakis, *Antenna Engineering Handbook*. New York, NY, USA: McGraw-Hill, 2007.

[35] M. Pinuela, P. D. Mitcheson, and S. Lucyszyn, "Analysis of scalable rectenna configurations for harvesting high frequency ambient radiation," in *Proc. Power MEMS*, Leuven, Belgium, 2010, pp. 41–44.

[36] J. O. McSpadden, L. Fan, and K. Chang, "Design and experiments of a high-conversion-efficiency 5.8-GHz rectenna," *EEE Trans. Microw. Theory Techn.*, vol. 46, no. 12, pp. 2053–2060, Dec. 1998.

[37] "HSMS-285x series datasheet," Avago Technol., San Jose, CA, USA, 2009.

[38] "Ultra low power boost converter with battery management for energy harvester applications," Texas Instruments Incorporated, Dallas, TX, USA, BQ25540 Datasheet, 2011.

[39] A. Nimo, D. Grgic, and L. M. Reindl, "Ambient electromagnetic wireless energy harvesting using multiband planar antenna," in *IEEE 9th Int. Syst., Signals, Devices Multi-Conf.*, Mar. 20–23, 2012, pp. 1, 6.

[40] H. Takhedmit, L. Cirio, B. Merabet, B. Allard, F. Costa, C. Vollaire, and O. Picon, "A 2.45-GHz dual-diode rectenna and rectenna arrays for wireless remote supply applications," *Int. J. Microw. Wireless Technol.*, vol. 3, no. 3, pp. 251–258, Jun. 2011.

[41] S. Vemuru, P. Singh, and M. Niamat, "Modeling impact of bypass diodes on photovoltaic cell performance under partial shading," in *IEEE Int. Electro/Information Technol. Conf.*, 2012, pp. 1–5.

[42] P. D. Mitcheson, E. M. Yeatman, G. K. Rao, A. S. Holmes, and T. C. Green, "Energy harvesting from human and machine motion for wireless electronic devices," *Proc. IEEE*, vol. 96, no. 9, pp. 1457–1486, Sep. 2008.

[43] R. J. M. Vullers, R. V. Schaijk, H. J. Visser, J. Penders, and C. V. Hoof, "Energy harvesting for autonomous wireless sensor networks," *IEEE Solid-State Circuits Mag.*, vol. 2, no. 2, pp. 29–38, Spring, 2010.

[44] V. Leonov, P. Fiorini, S. Sedky, T. Torfs, and C. Van Hoof, "Thermoelectric MEMS generators as a power supply for a body area network," in *13th Int. Solid-State Sens., Actuators, Microsyst. Conf. Tech. Dig.*, vol. 1, pp. 291–294.

Manuel Piñuela (M'12) received the BSc. degree in electrical and electronic engineering from the National Autonomous University of Mexico (UNAM), Mexico City, Mexico, in 2007, and is currently working toward the Ph.D. degree at Imperial College London, London, U.K.

From 2006 to 2009, he worked in industry as an Electronic Design and Project Engineer for companies focused on oil and gas services in both Mexico and the US. His research interests are wireless power transfer, RF power amplifiers, and RF energy harvesting.

Mr. Piñuela was the recipient of the Gabino Barrera Medal.

Paul D. Mitcheson (SM'12) received the M.Eng. degree in electrical and electronic engineering and Ph.D. degree from Imperial College London, U.K., in 2001 and 2005, respectively.

He is currently a Senior Lecturer with the Control and Power Research Group, Electrical and Electronic Engineering Department, Imperial College London. His research interests are energy harvesting, power electronics, and wireless power transfer. He is involved with providing power to applications in circumstances where batteries and cables are not suitable. His work has been sponsored by the European Commission, EPSRC and several companies.

Dr. Mitcheson is a fellow of the Higher Education Academy.

Stepan Lucyszyn (M'91–SM'04) received the Ph.D. degree in electronic engineering from King's College London (University of London), London, U.K., in 1992, and the D.Sc. (higher doctorate) degree from Imperial College London, London, U.K., in 2010.

He is currently a Reader (Associate Professor) of millimeter-wave electronics and Director of the Centre for Terahertz Science and Engineering, Imperial College London. After working in industry as a Satellite Systems Engineer for maritime and military communications, he spent the first 12 years researching microwave and millimeter-wave RF integrated circuits (RFICs)/monolithic microwave integrated circuits (MMICs), followed by RF microelectromechanical systems (MEMS) technologies. He has coauthored approximately 140 papers and 11 book chapters in applied physics and electronic engineering. He has delivered many invited presentations at international conferences.

Dr. Lucyszyn was an associate editor for the JOURNAL OF MICROELECTROMECHANICAL SYSTEMS (2005–2009). In 2011, he was the chairman of the 41st European Microwave Conference, Manchester, U.K. In 2005, he was elected Fellow of the Institution of Electrical Engineers (IEE), U.K., and Fellow of the Institute of Physics, U.K. In 2008, was invited as a Fellow of the Electromagnetics Academy, USA. In 2009, he became an IEEE Distinguished Microwave Lecturer (2010–2012). He is currently an Emeritus DML for 2013 and a newly appointed European Microwave Lecturer (EML) for the European Microwave Association.

cogent
engineering

CrossMark

ELECTRICAL & ELECTRONIC ENGINEERING | RESEARCH ARTICLE

A dual-band RF energy harvesting circuit using 4th order dual-band matching network

Sachin Agrawal[1]*, Manoj S. Parihar[1] and P.N. Kondekar[1]

*Corresponding author: Sachin Agrawal, PDPM-Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India
E-mail: bitssachin.agrawal@gmail.com

Abstract: A novel compact rectifier for dual-band operation in the RF energy harvesting is presented. The circuit comprises a 4th order dual-band impedance matching and a single-series circuit with one double diode, both are integrating into a compact shape to occupy a small area of 30×35 mm^2. The merit of the proposed rectifier circuit is that it can be extended to n number of the frequency band by using only $2 \times n$ matching elements. To validate the design method experimentally, a prototype of a dual-band rectifier is fabricated for two public telecommunication bands of GSM-900 and 1800. In order to reduce the circuit complexity and sensitivity arising due to lumped elements, the meander line and the open stub are used to realize the proposed circuit. A good agreement is obtained between the simulation and the measurement. The measured results show that the proposed rectifier circuit exhibits the conversion efficiency of 25.7 and 65% for an input power of −20 and 0 dBm, respectively. In addition, diode nonlinearity which affects the performance of the rectifier in terms of impedance matching is also investigated.

Subjects: Electromagnetics & Microwaves; Electronics; Circuits & Devices

Keywords: RF energy harvesting; dual band impedance matching; rectifier; RF-to-dc-conversion efficiency; frequency transformation

ABOUT THE AUTHOR

Sachin Agrawal did his BE and ME from Jiwaji University and Birla Institute of Technology Pilani, India in 2005 and 2009, respectively. Currently, he is pursuing his PhD from PDPM-Indian Institute of Information Technology Design and Manufacturing Jabalpur, India. His area of interest are; RF energy harvesting circuit, Antenna including planar and dielectric resonator antenna and Power amplifier.

Sachin Agrawal

PUBLIC INTEREST STATEMENT

With rapid growth in wireless communication, a huge amount of radio frequency (RF) energy broadcasted through billions of microwave sources such as mobile phones, handheld radios, and radio broadcast stations. Therefore, it is meaningful to collect and supply it to many electrical devices like mobile headsets, wearable medical sensors through RF energy harvesting. Since the ambient RF energy is distributed in multiple frequency bands, therefore the amount of energy harvested could increase if the circuit is designed for multiple frequency bands. In this work, we present a compact dual-band energy harvesting circuit to harvest energy from two most useful frequency bands, GSM-900 and 1800. The merit of the proposed rectifier circuit is it can be extended to n number of the frequency band by using only $2 \times n$ matching elements. A prototype is fabricated, and its performance is evaluated using Vector Network Analyzer (VNA). The total size of the rectifier is about 30×35 mm^2.

cogent·oa

cogent · engineering
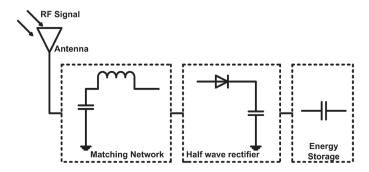
1. Introduction

A revolutionary growth in wireless technology attracts huge attention from research community to make the self-sustainable device feasible through RF energy harvesting. It exploits ambient electromagnetic energy transmitted from different RF systems to remotely feed the electronic devices (Nintanavongsa, Muncuk, Lewis, & Chowdhury, 2012). Compared to other harvesting techniques, RF energy harvesting provides relatively predictable energy supply owing to the features of easy availability and less dependency on environmental variations. The typical block diagram of RF energy harvesting circuit is shown in Figure 1. It consists of three major blocks viz; antenna, matching network (MN), diode detector followed by an energy storage. The first element, antenna is employed to capture the RF signals of different frequencies and polarization, while second MN is for maximum power transfer, and the last rectifier is used to convert the RF energy to dc voltage. It means harvesting circuit performance can be evaluated in terms of accessible ambient RF energy and its conversion rate (Agrawal, Pandey, Singh,& Parihar, 2014). These parameters are heavily influenced by surrounding terrain conditions as the multiple reflection and dissipation certainly deteriorate the level of available ambient RF energy. As a result, conversion efficiency and dc output voltage may degrade. Previously, the majority of available RF energy harvesting circuits focused on single frequency band hence offer low dc output voltage. As the multiple RF energy sources of different frequency bands are available, thus from an ambient RF harvesting perspective, the output dc voltage could be increased if the circuit is designed for multiple frequency bands rather than a single band. A wide-band energy harvester can also promise a high output voltage by accumulating the number of RF signals at a time. However, due to nonlinear behavior of the diode, harvesting circuit itself exhibits nonlinearity i.e. its input impedance varies with the received RF power. Thus, it is quite difficult to retain the impedance match and high conversion efficiency over a large frequency range (Song, Huang, Zhou, & Carter, 2014). The losses due to impedance mismatch over a large bandwidth can be illustrated in Collado and Georgiadis (2013), where only 8% conversion efficiency is achieved at -20 dBm.

To address this, it is preferable to harvest energy from several narrow frequency bands rather than a single large one. In literature, numerous topologies have been proposed to accomplish the multi-band energy harvesting (Bergès, Fadel, Oyhenart, Vigneras, & Taris, 2015; Hamano et al., 2016; Ho et al., 2016; Keyrouz, Visser, & Tijhuis, 2013; Kuhn, Lahuec, Seguin, & Person, 2015; Liu, Zhong, & Guo, 2015; Niotaki, Georgiadis, Collado, & Vardakas, 2014; Pinuela, Mitcheson, & Lucyszyn, 2013; Scheeler, Korhummel, & Popovic, 2014; Shariati, Rowe, Scott, & Ghorban, 2015; Sun, Guo, He, & Zhong, 2013). These topologies can be differentiated in terms of filter functionality i.e. how the antenna or source impedance is matched to the rectifier circuit. For instance, in Pinuela et al. (2013) and Keyrouz et al. (2013) several single-band rectennas (combination of antenna and rectifier circuit) were stacked to constitute a multi-band harvesting circuit. In this case, each rectenna was designed for a specific frequency band. Thus, for compact applications, this architecture is not suitable due to the number of antennas used. Moreover, in most of the reported works, the quality assessment of the output voltages combination was not taken into consideration. In Kuhn et al. (2015), the circuit complexity is reduced to a certain extent by replacing the multiple antennas with a single wide-band antenna. However, in this topology too, the number of rectifiers increases with the frequency bands, which leads to prolonging the circuit complexity.

Figure 1. Typical block diagram of RF energy harvesting circuit.

Besides, a multi-band harvesting circuit can also be formed by simply embedding a multi-band matching network between the multi-band antenna and the rectifying circuit (Bergès et al., 2015; Hamano et al., 2016; Ho et al., 2016; Liu et al., 2015; Niotaki et al., 2014; Scheeler et al., 2014; Shariati et al., 2015; Sun et al., 2013). The multi-band matching network can be designed either by distributed or by lumped element. In general, the multi-band rectifier circuit experiences two types of losses: first due to shift in resonance frequency from the optimum frequency point, and second due to the filter complexity. Because of the diode nonlinearity, the input impedance of the circuit varies as a function of power and frequency which causes a shift in resonance frequency. The difficulty due to diode nonlinearity can be observed in Sun et al. (2013) where the dual-band rectifier circuit exhibits impedance matching for a small range of input power. The losses induced because of filter complexity can be observed in the recently reported works on dual band harvesting circuit (Niotaki et al., 2014; Scheeler et al., 2014; Shariati et al., 2015). In Niotaki et al. (2014), for $P_{in} = -15$ dBm, author achieved the conversion efficiency of 23% at the expense of increased filter complexity consisting two series and two shunt pairs of reactive elements. Thus, for more than dual band applications, the proposed circuit topology is not suitable due to excessive filtering components used. To obtain good conversion efficiency a dual-band rectenna reported in Scheeler et al. (2014). However, the rectenna was large in size and requiring a complex impedance tuning circuit. In Shariati et al. (2015) also, a dual-band matching network consisting nine reactive elements was employed to achieve the dual-band characteristics.

In order to reduce the filter complexity, this work proposed a compact dual-band harvesting circuit for GSM-900 and 1800. It consists of a 4th order dual-band matching network based on $1 - n$ frequency transformation, which is optimized for the energy harvesting circuit to reduce the complexity up to $2 \times n$ reactive elements (n is the number of frequency bands). Similar to frequency transformation method, the proposed dual-band rectifier circuit can be extended to n number of frequency bands by using the $2 \times n$ number of reactive elements. The detailed analysis and design guidelines of dual band rectifier circuit are discussed in Section 2.

2. Dual band rectifier design and analysis

This section presents the design and analysis of a dual-band harvesting circuit in terms of impedance matching, DC output voltage and RF-to-dc conversion efficiency. The topology of the proposed dual-band RF energy harvesting circuit is shown in Figure 2(a). As seen, the low-cost Schottky diode is used to transform the input RF power to DC voltage. The impedance matching at two frequency is achieved using a series and parallel combination of the *LC* pair. The main idea underlying the suggested multi-band matching network is $1 - n$ frequency transformation (one to many mapping of frequency), which transforms a single-band matching network to multi-band matching network (Nallam & Chatterjee, 2013). As the name ($1 - n$) suggests that for designing a multi-band matching network, primarily a single-band matching network is required whose resonant frequency is dependent on the frequencies for which multi-band matching network proposed to designed.

Moreover, this frequency transformation method depends on the type of load impedance, whether it is series or parallel combination of *RC* or *RL*. Since the selected diode (HSMS-2852) has capacitive behavior throughout the frequency, it can be represented in a series or parallel combination of *R* and *C*. In the case of parallel *RC* load, the following equations are used to transform the single-band matching network into multi-band matching network.
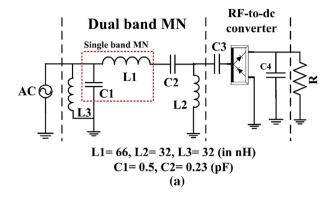
$$\omega = \frac{\omega_t^n + a_m \omega_t^{n-m} + a_{m+2} \omega_t^{n-(m+2)} + \cdots}{\omega_t^{n-1} + a_{m+1} \omega_t^{n-(m+1)} + a_{m+3} \omega_t^{n-(m+3)} + \cdots} \tag{1}$$
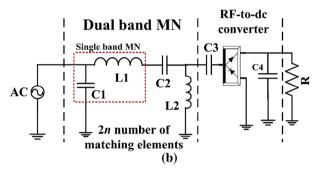
where, n is the number of bands and m varies from 2 to n. After substituting the value of n, Equation (1) can be expanded in partial fraction form using the causal foster analysis as:

$$\omega = \omega_t + \frac{1}{\frac{\omega_t}{a_2 - a_3} + \frac{1}{\cdots}} \tag{2}$$

Figure 2. (a) Circuit diagram of the proposed dual-band rectifier circuit and (b) optimized dual-band rectifier circuit.

The coefficients a_2 and a_3 can be calculated as:

$$\omega_m = \sum_{i=1}^{n} (-1)^{i-1} \omega_i \tag{3}$$

$$a_m = (-1)^m \sum_{i,j=1,1 \,\&\, i \neq j}^{n,n} (-1)^{i+j} \omega_i \omega_j \tag{4}$$

$$a_{m+1} = (-1)^{m+1} \sum_{i,j,k=1,1,1 \,\&\, i \neq j \neq k}^{n,n,n} (-1)^{i+j+k} \omega_i \omega_j \omega_k \tag{5}$$

$$a_{m+n} = (-1)^{m+n} \sum_{i,j,k\ldots=1,1,1\ldots i \neq j \neq k}^{n,n,n\ldots} (-1)^{i+j+k+\cdots} \omega_i \omega_j \omega_k \cdots \tag{6}$$
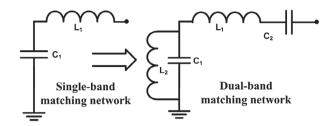
$$a_n = \prod_{i=1}^{n} (-1)^n \omega_i \tag{7}$$

Equation (6) is similar to that presented in Nallam and Chatterjee (2013), except the term $(-1)^{m+n}$, which is included here to realize the multi-band matching network for more than three frequency bands i.e. for $n \geq 3$.

   With this transformation, the capacitor of the matching network is transformed to the combination of prototype capacitor parallel with inductor whereas, an inductor is transformed into a combination of the same inductor with a series capacitor. Figure 3 shows the circuit schematic of transformation of a single-band matching network to the dual-band matching network. It can be seen that C_1 is transformed to $C_1 \| L_2$ and L_1 transformed to L_1 series with C_2. After successful usage of (1)–(7), the resultant multi-band matching network requires $3n - 1$ and $4n - 1$ reactive elements for L and Π-type topology, respectively.

Figure 3. Conversion of a single-band matching network to dual-band matching network using $1 - n$ frequency transformation.

As the aim is to design a dual-band harvesting circuit, therefore, we require here only 5 or 7 reactive elements with L and Π-type topologies, respectively. From Figure 2(a), it can be seen that the resultant matching network consists of five elements, where the inductor L_3 and capacitor C_2 results after the transformation of capacitor C_1 and inductor L_1, respectively. Besides, the inductor L_2 occurrs due to the diode reactive element, which is generally a capacitor.

In this work, two frequencies 0.9 and 1.8 GHz that correspond to the maximum signal strength are chosen for dual-band harvesting circuit. According to this method, it is necessary to assign the frequencies in descending order e.g. $\omega_1 = 1.8, \omega_2 = 0.9$. Therefore, from (3) single-band matching network frequency is equal to $\omega_1 - \omega_2 = 2\pi(1.8 - 0.9) \times 10^9 = 0.9 \times 2\pi \times 10^9$. In order to match the source impedance with the rectifier at the calculated frequency 0.9 GHz, the chosen matching topology is L-type as shown by the encircled portion in Figure 2(a). The corresponding element values can be approximated using the various methods some of which are described in Pozar (2010). Subsequently, this single-band matching network is transformed to dual-band matching network using (1)–(7). The detailed design steps of the dual-band rectifier circuit are summarized as follows:

(1) As we are interested in matching the diode to 50 Ω at two frequencies (0.9 and 1.8 GHz) so, the order of transformation is equal to 2 or $n = 2$.

(2) In the first step, single-band matching network is designed at the frequency f calculated as: $f = f_2 - f_1 = 1.8 - 0.9 = 0.9$ GHz. In this case, any matching topology that matches the diode to 50 Ω, at 0.9 GHz for an input power $P_{in} = -20$ dBm, and load resistance 4.7 kΩ can be used. The chosen single-band matching network is shown by the encircled portion in Figure 2(a).

(3) Afterwards, this single-band matching network is transformed into dual-band using the (1)–(7) as shown below:

Since $n = 2$ therefore, from (2)

$$\omega = \omega_t + \frac{1}{\frac{\omega_t}{a_2}} \tag{8}$$

From (4) a_2 can be calculated as:

$$a_2 = (-1)^{2+2} \sum (-1)^{1+2} \omega_1 \omega_2 = \omega_1 \omega_2 \tag{9}$$

$$a_2 = -1.62 \times 4\pi^2 \times 10^8 = 0.64 \times 10^{20} \tag{10}$$

Thus, inductor L_1 (=66 nH) is transformed to impedance as:

$$j66 \times 10^9 \omega = j66 \times 10^9 \omega_t + \frac{1}{j0.23 \times 10^{-12} \omega_t} \tag{11}$$

Similarly, capacitors (=0.5 pF) are transformed to the admittance as:

$$j5 \times 10^{-13} \omega = j5 \times 10^{-13} \omega_t + \frac{1}{j32 \times 10^{-9} \omega_t} \tag{12}$$

The circuit schematic of the dual-band harvesting circuit is shown in Figure 2(a). It can be seen that resultant matching network consists of five reactive elements according to $3n - 1$. In order to reduce the circuit complexity and sensitivity due to reactive elements, a parametric study has been carried out to eliminate the elements showing minimum influence on the circuit performance.

Figure 4 shows the simulated $|S_{11}|$ for the different combination of matching elements. The simulated results demonstrate that $|S_{11}|$ experiences maximum change when inductor L_2 and capacitor C_2 are removed from the circuit, whereas it remains almost unaffected when L_3 is not present in the circuit. Therefore, inductor L_3 can be extruded from the circuit and the resultant matching circuit requires only $2n$ and $3n$ reactive elements in place of $3n - 1$ and $4n - 1$ elements. In this way, for each topology, the proposed circuit reduces $n - 1$ elements compared to the conventional method. Figure 2(b) demonstrates the optimized circuit diagram of the dual-band rectifier. It can be observed that circuit requires large inductors value of 32 and 66 nH. Thus, it is quite difficult to realize the practical rectifier circuit whose response is similar to the response of simulated result. In order to avoid any impedance mismatch due to the small difference in elements value, the meander line inductor and open stub are used to realize the inductors and capacitors, respectively. In this case, not only fabrication and optimization process become so easy but the cost will also reduced.

Figure 5, shows the layout of the dual-band rectifier circuit. In Nintanavongsa et al. (2012) and Agrawal et al. (2014), it has been demonstrated that the number of rectifying diodes or equivalently voltage multiplier stages are very much sensitive to the RF-to-dc conversion efficiency. In low-power region (≤ -20 dBm), efficiency decreases if voltage multiplier stages increase, whereas in higher power region (≥ -20 dBm), an opposite effect occurs. As the demand is to harvest energy in low-power region, single-series circuit with a double diode is used to convert received RF energy into dc voltage. From the left side of the circuit, the first meander line corresponds to the inductor L_1, while the second meander line represents the inductor L_2, of the Figure 2(b). The shunt stub is accounted

Figure 4. Simulated $|S_{11}|$ vs. frequency for different combination of circuit elements.
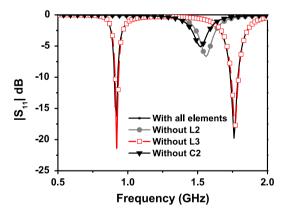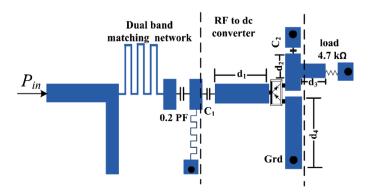
Figure 5. Layout of the proposed dual-band RF energy harvesting circuit.

for the shunt capacitor C_1 of Figure 2(b). The dimensions of each element are calculated according to their respective reactive element value and the substrate on which circuit has to be fabricated. In Assimonis, Daskalakis, and Bletsas (2016), it has been demonstrated that traces (microstrips) connected to the rectifier terminals (e.g. distance between the diode and via and diode and load) are highly sensitive for RF-to-dc efficiency. Therefore, traces d_1, d_2 between diode and capacitor C_1 and C_2, d_3 between diode and load and d_4 between diode and ground are adjusted to optimize the impedance matching as well as the conversion efficiency of the rectifier. Due to nonlinear behavior of the diode, harvesting circuit itself exhibits nonlinearity i.e. its input impedance varies with received RF power, therefore harmonic-balance (HB) and large signal analysis (LSSP) were employed to take into consideration the nonlinear behavior of the rectifier.

The photograph of the fabricated dual-band rectifier is shown in Figure 6. It is fabricated on a 1.54 mm thick FR-4 substrate with a dielectric constant (ϵ_r) of 4.3 using chemical etching method. The rectifier performance is evaluated in terms of $|S_{11}|$ and output voltage using the Agilent vector network analyzer (VNA). The simulated and measured $|S_{11}|$ is illustrated in Figure 7(a). The measured result shows reasonable agreement with the simulated one; the slight difference can be accounted for the fabrication imperfections. It is well known that impedance matching is a function of frequency and input power, due to the nonlinearity of the diode. Such a characteristic is examined in Figure 7(b), where the measured $|S_{11}|$ is demonstrated as a function of input power level for three different load impedance values. From results, it is clear that impedance matching of the harvesting circuit is greatly affected by the input power and the load impedance. Figure 7(b) demonstrates that as power increases, the impedance matching at 0.9 GHz degraded drastically, while at 1.8 GHz, it improves. Moreover, it is noticed that the impedance matching at higher power level is more sensitive to the variation of load impedance (*RL*).

The measured RF-to-dc conversion efficiency and output voltage vs. input power for both frequencies are demonstrated in Figure 8(a). For 0.9 GHz, efficiency is equal to 25.7 and 65.1% for an input power of −20 and 0 dBm, respectively. However, at 1.8 GHz the efficiency is relatively small that might be due to the increased parasitic losses in the rectifier diode. Figure 8(b) shows the relation between the output voltage and frequency for various input power levels at fixed load resistance value of 4.7 kΩ. It can be seen that maximum output voltage is achieved in the frequency range of 860–900 and 1770–1800 MHz, showing the rectifier's capability to harvest RF energy in the GSM-900 and 1800 bands.

Figure 9, depicts the measured conversion for various values of load resistance. It can be noticed that the circuit yields maximum efficiency when the load impedance is 4.7 kΩ. It starts decreasing as the load impedance varies from 4.7 kΩ. Table 1 shows the comparison of the conversion efficiency and the size of the proposed rectifier with the similar works reported previously. Only measured results are compared in Table 1. It can be seen that in low-power condition maximum efficiency is achieved in Sun et al. (2013), but the expense of bulky circuit size. However, the proposed rectifier offers an optimal conversion efficiency with compact circuit size.

Figure 6. Photograph of the fabricated dual-band RF energy harvesting circuit.

cogent · engineering
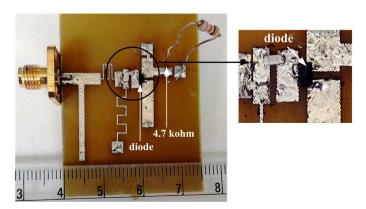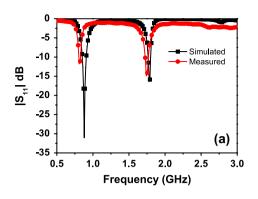
Figure 7. (a) Simulated and measured $|S_{11}|$ vs. frequency and (b) measured $|S_{11}|$ for various power levels.
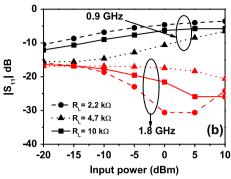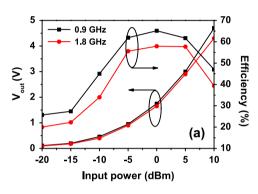
Figure 8. (a) Measured output dc voltage and RF-to-dc conversion efficiency and (b) measured dc voltage vs. frequency.
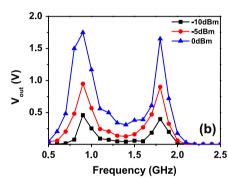
Figure 9. Measured RF-to-dc conversion efficiency for different load impedance.
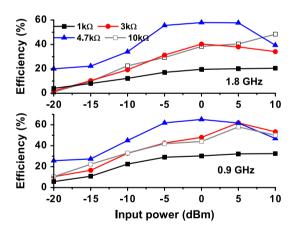
Table 1. Performance comparison of the proposed dual rectifier with recently published works			
Ref.	**Measured rectifier efficiency (%)**	**Input power (dBm)**	**Rectifier size**
Ho et al. (2016)	15.8 @ 0.89 GHz	−20	100×65 mm^2
	11.2 @ 1.76 GHz		
Hamano et al. (2016)	10 @ 2.15 GHz	−10	37×71 mm^2
	15 @ 5.84 GHz		
Bergès et al. (2015)	27 @ 0.91/2.4 GHz	−16	78×88 mm^2
Sun et al. (2013)	30 @ 2.14 GHz	−20	145 mm
	35 @ 1.84 GHz		
Liu et al. (15)	20 @ (0.91+1.8) GHz	−20	23×37 mm^2
This work	27.5 @ 0.9 GHz	−20	30×35 mm^2
	20 @ 1.8 GHz		

cogent ·· engineering

3. Conclusion

A new compact 4th order dual-band rectifier has been designed to harvest the RF power of GSM-900 and 1800 bands. In order to reduce the circuit complexity and sensitivity due to reactive elements, the meander line and the open stub are used to fabricate the matching network. For $P_{in} = -20$ dBm, the measured RF-to-dc conversion efficiency of 27.5 and 20% is achieved at 0.9 and 1.8 GHz, respectively. Further, more than 45 and 34% conversion efficiency is maintained from -10 to 10 dBm for 0.9 and 1.8 GHz, respectively.

Author details
Sachin Agrawal[1]
E-mail: bitssachin.agrawal@gmail.com
Manoj S. Parihar[1]
E-mail: mparihar@iiitdmj.ac.in
P.N. Kondekar[1]
E-mail: pnkondekar@iiitdmj.ac.in
[1] PDPM-Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India.

Citation information
Cite this article as: A dual-band RF energy harvesting circuit using 4th order dual-band matching network, Sachin Agrawal, Manoj S. Parihar & P.N. Kondekar, *Cogent Engineering* (2017), 4: 1332705.

References
Agrawal, S., Pandey, S. K., Singh, J., & Parihar, M. S. (2014). *Realization of efficient RF energy harvesting circuits employing different matching technique* (pp. 754–761). Fifteenth International Symposium on Quality Electronic Design, Santa Clara, CA.
Assimonis, S. D., Daskalakis, S. N., & Bletsas, A. (2016). Sensitive and efficient rf harvesting supply for batteryless backscatter sensor networks. *IEEE Transactions on Microwave Theory and Techniques, 64*, 1327–1338.
Bergès, R., Fadel, L., Oyhenart, L., Vigneras, V., & Taris, T. (2015). A dual band 915MHz/2.44 GHz RF energy harvester. *Microwave Conference (EuMC), 2015 European*. IEEE.
Collado, A., & Georgiadis, A. (2013). Conformal hybrid solar and electromagnetic (EM) energy harvesting rectenna. *IEEE Transactions on Circuits and Systems I: Regular Papers, 60*, p. 10.
Hamano, K., Tanaka, R., Yoshida, S., Miyachi, A., Nishikawa, K., & Kawasaki, S. (2016). Design of dual-band rectifier using microstrip spurline notch filter. *Radio-Frequency Integration Technology (RFIT), 2016 IEEE International Symposium on*. IEEE.
Ho, D. K., Kharrat, I., Ngo, V. D., Vuong, T. P., Nguyen, Q. C., & Le, M. T. (2016). *Dual-band rectenna for ambient RF energy harvesting at GSM 900 MHz and 1800 MHz* (pp. 306–310). 2016 IEEE International Conference on Sustainable Energy Technologies (ICSET), Hanoi.

Keyrouz, S., Visser, H., & Tijhuis, A. (2013, April). Multi-band simultaneous radio frequency energy harvesting. *In Antennas and Propagation (EuCAP), 2013 7th European Conference on*, 3058–3061.
Kuhn, V., Lahuec, C., Seguin, F., & Person, C. (2015). A multi-band stacked RF energy harvester with RF-to-DC efficiency up to 84\%. *IEEE Transactions on Microwave Theory and Techniques, 63*, 1768–1778.
Liu, Z., Zhong, Z., & Guo, Y.-X. (2015). Enhanced dual-band ambient RF energy harvesting with ultra-wide power range. *IEEE Microwave and Wireless Components Letters, 25*, 630–632.
Nallam, N., & Chatterjee, S. (2013, June). Multi-band frequency transformations, matching networks and amplifiers. *Circuits and Systems I: Regular Papers, IEEE Transactions on, 60*, 1635–1647.
Nintanavongsa, P., Muncuk, U., Lewis, D., & Chowdhury, K. (2012, March). Design optimization and implementation for rf energy harvesting circuits. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2*, 24–33.
Niotaki, K., Georgiadis, A., Collado, A., & Vardakas, J. S. (2014). Dual-band resistance compression networks for improved rectifier performance. *IEEE Transactions on Microwave Theory and Techniques, 62*, 3512–3521.
Pinuela, M., Mitcheson, P., & Lucyszyn, S. (2013, July). Ambient rf energy harvesting in urban and semi-urban environments. *Microwave Theory and Techniques, IEEE Transactions on, 61*, 2715–2726.
Pozar, D. M. (2010). *Microwave engineering* (3rd ed.). New Delhi: Wiley.
Scheeler, R., Korhummel, S., & Popovic, Z. (2014). A dual-frequency ultralowpower efficient 0.5-g rectenna. *IEEE Microwave Magazine, 15*, 109–114.
Shariati, N., Rowe, W. S., Scott, J. R., & Ghorban, K. (2015). Multi-service highly sensitive rectifier for enhanced RF energy scavenging. *Scientific Reports, 5*, 9655
Song, C., Huang, Y., Zhou, J., & Carter, P. (2014). Improved ultrawideband rectennas using hybrid resistance compression technique. *IEEE Transactions on Antennas and Propagation, 65*, 2057–2062.
Sun, H., Guo, Y. X., He, M., & Zhong, Z. (2013). A dual-band rectenna using broadband Yagi antenna array for ambient RF power harvesting. *IEEE Antennas and Wireless Propagation Letters, 12*, 918–921. doi:10.1109/LAWP.2013.2272873.

Cogent Engineering (ISSN: 2331-1916) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

Scientific Research

Optimization of the Voltage Doubler Stages in an RF-DC Convertor Module for Energy Harvesting

Kavuri Kasi Annapurna Devi[1*], Norashidah Md. Din[2], Chandan Kumar Chakrabarty[2]

[1]Department of Electrical and Electronic Engineering, INTI International University, Nilai, Malaysia
[2]Department of Electronics and Communication Engineering, Universiti Tenega Nasional, Kajang, Malaysia
Email: {[*]kavurik.adevi., chandan}@newinti.edu.my, norashidah@uniten.edu.my

ABSTRACT

This paper presents an optimization of the voltage doubler stages in an energy conversion module for Radio Frequency (RF) energy harvesting system at 900 MHz band. The function of the energy conversion module is to convert the (RF) signals into direct-current (DC) voltage at the given frequency band to power the low power devices/circuits. The design is based on the Villard voltage doubler circuit. A 7 stage Schottky diode voltage doubler circuit is designed, modeled, simulated, fabricated and tested in this work. Multisim was used for the modeling and simulation work. Simulation and measurement were carried out for various input power levels at the specified frequency band. For an equivalent incident signal of –40 dBm, the circuit can produce 3 mV across a 100 kΩ load. The results also show that there is a multiplication factor of 22 at 0 dBm and produces DC output voltage of 5.0 V in measurement. This voltage can be used to power low power sensors in sensor networks ultimately in place of batteries.

Keywords: Energy Conversion; RF; Schottky Diode; Villard; Energy Harvesting

1. Introduction

RF energy harvesting is one type of energy harvesting that can be potentially harvested such as solar, vibration and wind. The RF energy harvesting uses the idea of capturing transmitted RF energy at ambient and either using it directly to power a low power circuit or storing it for later use. The concept needs an efficient antenna along with a circuit capable of converting RF signals to DC voltage. The efficiency of an antenna mainly depends on its impedance and the impedance of the energy converting circuit. If the two impedances aren't matched then it will be unable to receive all the available power from the free space at the desired frequency band. Matching of the impedances means that the impedance of the antenna is the complex conjugate of the impedance of the circuit (voltage doubler circuit).

The concept of energy harvesting system is shown in **Figure 1**, which consists of matching network, RF-DC conversion and load circuits. The authors in [1], used a 2.4 GHz operating frequency with an integrated zero bias detector circuit using BiCMOS technology which produced an output voltage of 1 V into a 1 MΩ load at an input power level of 0 dBm. H. Yan and co-authors revealed that a DC voltage of 0.8 volts can be achieved from a –20 dBm RF input signal at 868.3 MHz through

simulation results [2]. In [3], work was carried out on a firm frequency of 900 MHz by matching to a 50 Ω impedance and resonance circuit transformation in front of the Schottky diode which yields an output voltage of over 300 mV at an input power level of 2.5 μ. W. J. Wang, L. Dong and Y. Fu [4] used a Cockcroft-Walton multiplier circuit that produced a voltage level of 1.0 V into a 200 MΩ load for an input power level of less than –30 dBm at a fixed frequency of 2.4 GHz.

The energy conversion module designed in this paper is based on a voltage doubler circuit which can be able to output a DC voltage typically larger than a simple diode rectifier circuit as in [5], in which switched capacitor charge pump circuits are used to design two phase voltage doubler and a multiphase voltage doubler. The module presented in this can function as an AC to DC converter that not only rectifies the input AC signal but also elevates the DC voltage level. The output voltage of the

Figure 1. Schematic view of a RF energy harvesting system.

CS

energy conversion module can be used to energize the low power devices for example sensors for a sensor network in application to agriculture.

Section 2 of this paper discusses on the theoretical background of the voltage doubler circuit. Section 3 presents the simulation study and implementation of the circuit design. Section 4 provides the results and analysis. Section 5 concludes with a discussion on the findings from the simulated and measured results.

2. Voltage Multiplier

There are various voltage multiplier circuit topologies. The design used in this module is derived from the function of peak detector or a half wave peak rectifier. The Villard voltage multiplier circuit was chosen in the circuit design of this paper because it produces two times of the input signal voltage towards ground at a single output and can be cascaded to form a voltage multiplier with an arbitrary output voltage and its design simplicity.

2.1. Diode Modeling

The voltage multiplier circuit in this design uses zero bias Schottky diode HSMS-2850 from Agilent. The attractive feature of these Schottky diodes are low substrate losses and very fast switching but leads to a fabrication overhead. This diode has been modeled for the energy harvesting circuit which comes in a one-diode configuration. The modeling parameters for these diodes are given by Agilent in their data sheets. These parameters are used in Multisim for its own modeling purposes. The modeling is done by transforming the diode into an equivalent circuit using passive components which are described by the SPICE parameters in **Table 1** [6].

The diode used in this design is shown in **Figure 2** and its equivalent model is shown in **Figure 3**. The special features of HSMS-2850 diode is that it provides a low forward voltage, low substrate leakage and uses the non

Figure 2. Schottky diode.

Figure 3. Linear circuit model of the Schottky diode [6].

symmetric properties of a diode that allows unidirectional flow of current under ideal condition.

The diodes are fixed and are not subject of optimization or tuning. This is described using the following derivations. By neglecting the effect of diode substrate, an equivalent linear model that can be used for the diode as shown in **Figure 3**. When C_j is the junction capacitance and R_j is the junction resistance, the admittance Y_z of the linear model is given by

$$Y_Z = Y_{C_j} + Y_{R_j} \tag{1}$$

Equation (1) related to the frequency of operation is given by

$$Y_Z = jwC_j + \frac{1}{R_j} \tag{2}$$

$$= \frac{jwC_jR_j + 1}{R_j} \tag{3}$$

The impedance Z of the linear model is given by

$$Z = \frac{R_j}{1 + jwR_jC_j} \tag{4}$$

The total impedance Z_T is given by

$$Z_T = R_S + \frac{R_j}{1 + jwR_jC_j} \tag{5}$$

where R_S is the series resistance of the circuit and R_j is given by

$$R_j = \frac{8.33 \times 10^{-5} \times N \times T}{I_b + I_S}$$

where:

I_b = bias current in μA;
I_s = saturation current in μA;
T = temperature (K);

Table 1. SPICE parameters.

Parameters	Units	HSMS 2850
B_V	V	3.8
C_{J0}	pF	0.18
E_G	Ev	0.69
I_{BV}	A	3E–4
I_S	A	3E–6
N	No unite	1.06
R_S	Ω	25
$P_B (V_J)$	V	0.35
$P_T (XTI)$	No units	2
M	No units	0.5

N = ideality factor.

In Equation (5), R_j and C_j are constants and the frequency of operation (w) is the only variable parameter. As the frequency increases, the value of Z is almost negligible compared to the series resistance R_S of the diode. From this it is concluded that the function of the diode is independent of the frequency of operation.

2.2. Single Stage Voltage Multiplier

Figure 4 represents a single stage voltage multiplier circuit. The circuit is also called as a voltage doubler because in theory, the voltage that is arrived on the output is approximately twice that at the input. The circuit consists of two sections; each comprises a diode and a capacitor for rectification. The RF input signal is rectified in the positive half of the input cycle, followed by the negative half of the input cycle. But, the voltage stored on the input capacitor during one half cycle is transferred to the output capacitor during the next half cycle of the input signal. Thus, the voltage on output capacitor is roughly two times the peak voltage of the RF source minus the turn-on voltage of the diode.

The most interesting feature of this circuit is that when these stages are connected in series. This method behaves akin to the principle of stacking batteries in series to get more voltage at the output. The output of the first stage is not exactly pure DC voltage and it is basically an AC signal with a DC offset voltage. This is equivalent to a DC signal superimposed by ripple content. Due to this distinctive feature, succeeding stages in the circuit can get more voltage than the preceding stages. If a second stage is added on top of the first multiplier circuit, the only waveform that the second stage receives is the noise of the first stage. This noise is then doubled and added to the DC voltage of the first stage. Therefore, the more stages that are added, theoretically, more voltage will come from the system regardless of the input. Each independent stage with its dedicated voltage doubler circuit can be seen as a single battery with open circuit output voltage V_0, internal resistance R_0 with load resistance R_L, the output voltage, V_{out} is expressed as in Equation (7).

Figure 4. Single stage voltage multiplier circuit [7].

$$V_{out} = \frac{V_0}{R_0 + R_L} R_L \tag{6}$$

When n number of these circuits are put in series and connected to a load of R_L in Equation (6) the output voltage V_{out} obtained is given by this change in RC value will make the time constant longer which in turn retains the multiplication effect of two in this design of seven stage voltage doubler.

$$V_{out} = \frac{nV_0}{nR_0 + R_L} = V_0 \frac{1}{\frac{R_0}{R_L} + \frac{1}{n}} \tag{7}$$

The number of stages in the system has the greatest effect on the DC output voltage, as shown from Equations (6) and (7).

It is inferred that the output voltage V_{out} is determined by the addition of R_0/R_L and $1/n$, if V_0 is fixed. From this analysis it is observed that V_0, R_0 and R_L are all constants. Assume that $V_0 = 1$ V, $R_0/R_L = 0.25$, $n = 2, 3, 4, 5, 6$ and 7, the output voltage $V_{out} = 1.33$ V, 1.72 V, 2.0 V, 2.22 V, 2.43 V and 2.56 V respectively when substituted analytically in the Equation (7). This analysis can be compared with the results obtained in the circuit design of this module. In simulation at $n = 4$, $V_{out} = 1.42$ V, $n = 5$, $V_{out} = 1.67$ V; $n = 6$, $V_{out} = 1.92$; $n = 7$, $V_{out} = 2.15$ V; $n = 8$, $V_{out} = 1.92$ V; $n = 9$, $V_{out} = 1.81$ V. Also in measurement, for $n = 4$, $V_{out} = 2.1$ V; $n = 5$, $V_{out} = 2.9$ V; $n = 6$, $V_{out} = 3.72$ V; and $n = 7$, $V_{out} = 5$ V. As n increases, the increase in output voltage will be almost double the input voltage up to some number of stages. But at some point, *i.e.* beyond seven stages, in this circuit the output voltage gained (8 and 9 stages) will be negligible as shown in **Figure 5**.

The capacitors are charged to the peak value of the input RF signal and discharge to the series resistance (R_s) of the diode. Thus the output voltage across the capacitor of the first stage is approximately twice that of the input signal. As the signal swings from one stage to other, there is an additive resistance in the discharge path of the

Figure 5. Normalized output voltage multiplier versus number of stages.

diode and increase of capacitance due to the stage capacitors.

2.3. Seven Stage Voltage Multiplier

The seven stage voltage multiplier circuit design implemented in this paper is shown in **Figure 6**. Starting on the left side, there is a RF signal source for the circuit followed by the first stage of the voltage multiplier circuit. Each stage is stacked onto the previous stage as shown in the **Figure 6**. Stacking was done from left to right for simplicity instead of conventional stacking from bottom to top.

The circuit uses eight zero bias Schottky surface-mount Agilent HSMS-285X series, HSMS-2850 diodes. The special features of these diode is that, it provides a low forward voltage, low substrate leakage and uses the nonsymmetric properties of a diode that allows unidirectional flow of current under ideal conditions. The diodes are fixed and are not subject of optimization or tuning. This type of multiplier produces a DC voltage which depends on the incident RF voltage. Input to the circuit is a predefined RF source. The voltage conversion can be effective only if the input voltage is higher than the Schottky forward voltage.

The other components associated with the circuit are the stage capacitors. The chosen capacitors for this circuit are of through-hole type, which make it easier to modify for optimization, where in [8] the optimization was accomplished at the input impedance of the CMOS chip for a three stage voltage multiplier. The circuit design in this paper uses a capacitor across the load to store and provide DC leveling of the output voltage and its value only affects the speed of the transient response. Without a capacitor across the load, the output is not a good DC signal, but more of an offset AC signal.

In addition to the above, an equivalent load resistor is connected at the final node. The output voltage across the load decreases during the negative half cycle of the AC input signal. The voltage decreases is inversely proportional to the product of resistance and capacitance across the load. Without the load resistor on the circuit, the voltage would be hold indefinitely on the capacitor and look like a DC signal, assuming ideal components. In the design, the individual components of the stages need not to be rated to withstand the entire output voltage. Each component only needs to be concerned with the relative voltage differences directly across its own terminals and of the components immediately adjacent to it. In this type of circuitry, the circuit does not change the output voltage but increases the possible output current by a factor of two. The number of stages in the system is directly proportional to the amount of voltage obtained and has the greatest effect on the output voltage as explained in the Equation (7) and shown in **Figure 5**.

3. Simulation and Implementation

Multisim software was chosen for modeling and simulation which is a circuit simulation tool by Texas Instruments. The simulation and practical implementation were carried out with fixed RF at 945 MHz ± 100 MHz, which are close to the down link center radio frequency (947.5 MHz) of the GSM-900 transmitter. The voltages obtained at the final node V_{DC} of the multiplier circuit were recorded for various input power levels from −40 dBm - +5 dBm with power level interval (spacing) of 5 dBm.

The simulations were also carried out using same stage capacitance value (3.3 nF) and then with a varied capacitance value for all stages from 4 stages through 9 stages [9]. The capacitance value was varied in such a way that, from one stage to the next, it was halved. For example, if the first stage was 3.3 nF, the second stage was 1.65 nF, third stage was 825 pF, fourth stage was 415 pF and so on. But keeping in view of testing, the capacitance values were chosen to have a close match with the standard available values in the market.

Simulation was carried out through 4 to 9 voltage doubler stages. Based on results obtained a 7 stage doubler is best to implemented for this application.

The design of the printed circuit board (PCB) was carried out using DipTrace software. The material used to

Figure 6. Schematic of 7 stage voltage multiplier.

manufacture the printed circuit board (PCB) is the standard Fiberglass Reinforced Epoxy (FR4), with the thickness of 1.6 mm and dielectric constant of 3.9. The topology is constructed on the PCB with the dimensions of 98 mm × 34 mm (W × H). The Sub Miniature version A (SMA) connectors are used at the input and output of PCB to carry out the measurements. The circuit components consist of active and passive components. The component used in circuit is shown in **Table 2**.

Special handling precautions have been taken to avoid Electro Static Discharge (ESD), while assembling of the surface-mount zero bias Schottky diodes. Also special attention has been given to mount other components and the SMA connectors on to the PCB. The Photograph of Assembled circuit board I shown in **Figure 7**.

4. Results and Analysis

The simulated and measured results at the output voltage of voltage multiplier circuit are shown graphically in **Figure 8**. From the graph analysis, the simulated and the measured results agree considerably with each other. The measured results are shown to be better than the simulation results. The reason behind this may be due to the uncertainty in series resistance value of the diode obtained from SPICE parameters in modeling as explained in Equation (5). This resistance vale of diodes in practical circuit may be lower than in the model, which provides fast discharge path, in turn rise in voltage as passes through the stages and reaches to final output. In this work, the DC output voltages obtained through simulation and measurement at 0 dBm re 2.12 V and 5.0 V respectively. These results are comparatively much better than in ref. [9], where in at 0 dBm, 900 MHz they achieved 0.5 V and 0.8 V through simulation and measurement

respectively.

Figures 9 and **10** show the result of a 4 stage voltage doubler circuit with equal and varied capacitance values between the stages as described in Section 3.

From the analysis of these two simulations, it can be observed that the resulting output voltages are equal. The only difference between these two graphs is the rise time of the circuit with varied capacitance value is a little bit slower. But, overall result on the performance of rise time is still under 20 μs to 24 μs and the difference is negligible. From these results, the use of equal stage capacitance of each being 3.3 nF was hence considered for the design of the multiplier.

The results from **Figure 11**, shows that the output voltage reaches to 1.0 V within 20 μS and then uniformly increasing to 1.4 V, 1.67 V, 1.87 V and 2.12 V for 4, 5, 6 and 7 stages respectively compared to 2 mS as shown in [10]. **Figure 12** shows that the conversion ratio of 22 is achieved at 0 dBm input power and drops to 2.5 at –40 dBm. The highest value at 0 dBm is due to the innate characteristics of the zero bias Schottky diodes which conduct fairly well at higher input voltages.

5. Conclusion

From the experimental results, it is found that the pro-

Table 2. Component used in 7 stage voltage multiplier.

Name of component	Label	Value
Stage capacitors	C_1 - C_{14}	3.3 nF
Stage diodes	D_1 - D_{14}	HSMS 2850
Filter capacitor	C_L	100 nF
Load resister	R_L	100 kΩ

Input SMA Connector   Schottky Diodes   Stage Capacitors   Output SMA Connector

Filter Circuit

Figure 7. Photograph of assembled circuit board.

Figure 8. Simulated and test DC output voltage of multiplier as a function of input power.

Figure 9. DC output voltage verses rise time of 4 stage voltage doubler circuit with equal stage capacitance [8].

Figure 10. DC output voltage verses rise time of 4 stage voltage doubler with varied stage capacitance [8].

Figure 11. DC output voltage verses rise time of voltage doubler circuit through 4 - 7 stages with equal stage capacitance.

Figure 12. Conversion ratio as a function of input power.

posed voltage multiplier circuit operates at the frequency of 945 MHz with the specified input power levels. The results have shown that there is multiplication of the input voltage. From **Figure 12**, it is shown that at 0 dBm input power, the multiplication factor is 22. This is sig-

nificant, as the work shows that RF energy in the GSM-900 band can be harvested from the ambient RF source using the Villard circuit topology. The power density levels from a GSM base station is expected from 0.1 mW/m^2 to 1 mW/m^2 for a distance ranging from 25 m - 100 m. These power levels may be elevated by a factor between one and three for the GSM-900 downlink frequency bands depending on the traffic density [10]. The next phase of the research work is to interface the voltage multiplier circuit through a matching network to the antenna at the input side and a low power device to power from the system at the output side to complete the RF energy harvesting system.

6. Acknowledgements

REFERENCES

[1] S. von der Mark and G. Boeck, "Ultra Low Power Wakeup Detector for Sensor Networks," *SBMO/IEEE MTT-S International Microwave & Optoelectronics Conference* (*IMOC* 2007), Berlin, 29 October-1 November 2007, pp. 865-868. doi:10.1109/IMOC.2007.4404394

[2] H. Yan, J. G. M. Montero, A. Akhnoukh, L. C. N. de Vreede and J. N. Burghartz, "An Integration Scheme for RF Power Harvesting," *Proceedings of STW Annual Workshop on Semiconductor Advances for Future Electronics and Sensors*, Veldhoven, November 2005, pp. 64-66. doi:10.1.1.126.5786

[3] T. Ungan and L. M. Reindl, "Concept for Harvesting Low Ambient RF-Sources for Microsystems," 2007. http://www.imtek.de/content/pdf/public/2007/powermems_2007_paper_ungan.pdf

[4] J. Wang, L. Dong and Y. Z. Fu, "Modeling of UHF Voltage Multiplier for Radio-Triggered Wake-Up Circuits," *International Journal of Circuit Theory and Application*, Vol. 39, No. 11, 2010, pp. 1189-1197. doi:10.1002/cta.692

[5] J. A. Starzyk, Y.-W. Jan and F. J. Qiu, "A DC-DC Charge Pump Design Based on Voltage Doublers," *IEEE Transactions on Circuits and Systems-I*: *Fundamental Theory and Applications*, Vol. 48, No. 3, 2001, pp. 350-359.

[6] HSMS-2850, "Surface Mount Zero Bias Schottky Detector Diodes." http://www.crystal-radio.eu/hsms285xdata.pdf

[7] D. W. Harrist, "Wireless Battery Charging System Using Radio Frequency Energy Harvesting," M.S. Thesis, University of Pittsburgh, Pittsburgh, 2004.

[8] E. Bergeret, J. Gaubert, P. Pannie and J. M. Gaultierr, "Modeling and Design of CMOS UHF Voltage Multiplier for RFID in an EEPROM Compatible Process," *IEEE Transactions on Circuits and Systems-I*, Vol. 54, No. 10, 2007, pp. 833-837. doi:10.1109/TCSII.2007.902222

[9] B. Emmanuel, J. Gaubert, P. Pannier and J. M. Gaultier, "Conception of UHF Voltage Multiplier for RFID Circui," *IEEE North-East Workshop on Circuits and Systems*, Gatineau, June 2006, pp. 217-220. doi:10.1109/NEWCAS.2006.250961

[10] H. J. Visser, A. C. F. Reniers and J. A. C. Theeuwes, "Ambient RF Energy Scavenging: GSM and WLAN Power Density Measurements," *Proceedings of the* 38*th European Microwave Conference*, Eindhoven, 27-31 October 2008, pp. 721-724. doi:10.1109/EUMC.2008.4751554

AC0C

RF Current Meter

Stop Guessing and Built This Simple Meter Making Measurements Easy to Take

Why a Hand Held RF Current Meter

There are two things that I wanted to do which would be made easy with a current meter... Both are related to the unusual nature of the antenna here at AC0C.

1. Checking common mode surface current on coax and control cables up in the attic antennas.

With the high cost of ferrites, placing them with the aid of a current meter is very helpful in that you can compare the before (no ferrite) and after (with ferrites attached) - and know that you have addressed the current problem on a given bit of line.

Otherwise, it seemed to me, that ferrite placement was more of a guess than a science.

2. Measuring relative antenna currents along an element

With the W8WWV RVM system, I can measure element currents where ever a current sensor is placed. This is typically at the center of a dipole element.
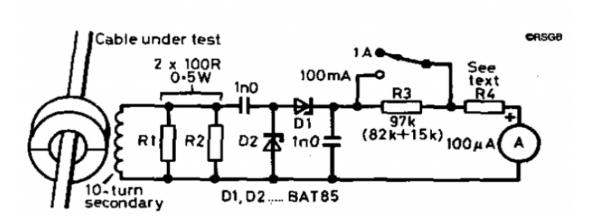
However, the center may or may not be the actual high-current point for the element depending on the frequency. My EZNEC models can indicate where the current peaks are - and by looking along these lines manually - comparing them with readings taken from the driven element, I can make current checks even in non-standard location of the array.

The meter described below accomplishes these two tasks very nicely.

Design Idea

The design follows an old RSGB design that I found on G3SEK's excellent web site:

http://www.ifwtech.co.uk/g3sek/clamp-on/clamp-on.htm

I followed this design pretty close except for the R3/R4 which I used variable trim pots. Due to the voltage-doubler nature of the circuit, it's capable of exceptional sensitivity.

The meter was calibrated for full-scale @ 1A on the higher scale, but I have left the lower scale set on a very high sensitivity setting which I find very handy for the common mode checking. Here, a sensitive meter providing a relative

current indication is needed.

Construction Details

Once parts were in hand, construction was very easy and took perhaps an hour.

No attempt was made to beautify the work - it's 100% orientated toward functionality and utility.

I had a 100 uA meter in the junk box along with the other parts needed and orienting them on the back-side of the meter was really the most time consuming aspect.

The switch is held in position with hot-glue. And a bit of hot glue is applied to the meter terminations as a safety precaution - just in case I were to brush the meter up against something carrying high levels of current.

Wraps around the toroid were made with wire-wrap type wire, 30 Gage.

Selection of the toroid used was based on what would physically fit around the RG-213 sized cable and could be easily opened and closed.
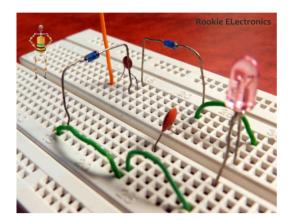
The rubber band shown in the picture makes for a very convenient closing mechanism - and in a lot of applications, I simply hold the toroid closed with my finger pressure which makes moving the meter along a wire - and moving from cable to cable - very quick and easy.
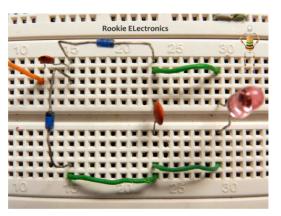
For more ideas, check out Frank N4SPP's very nice RF current meter found HERE.
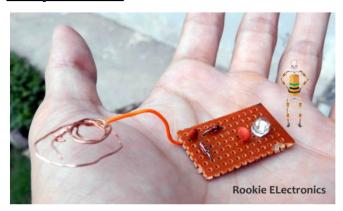
Created with the QTH.com SiteBuilder.

Bread board Arrangement:

Strip Board:

RF Diagnostics, LLC

WELCOME TO RF DIAGNOSTICS
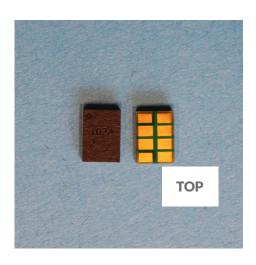
Energy Harvesting Modules

WIRELESS ENERGY HARVESTING PRODUCTS RF Diagnostics, LLC designs and sells surface mount RF to DC converters and other building block modules for energy harvesting systems. With our products you can harvest ambient AC/AM/FM/Wireless Energy to power your IoT battery free system. We also provide custom designs for RF/microwave and general electrical engineering products and design contracting services to clients in the commercial wireless and defense electronics businesses. We strive to be a high quality and cost effective design resource for your design team.

OUR PRODUCTS

TOP

RFD102A-DET: Microwave Energy Detector	RFD102A-TB: 60Hz...6GHz Energy Harvesting Test Board	RFD102A: Wireless Energy Harvesting Module
$18.00	$33.00	$12.00
Add to cart	Add to cart	Add to cart

OUR DESIGN SERVICES

Energy Harvesting Module & System Design

Any of our standard products can be customized for your application. We can support your application with matching, testing, antenna design, assembly and many other tasks. Have a new idea for a product? We can help you realize it. We are developing new products and are always interested in your feedback to guide our product development.

MMIC Design

We use Agilent's Advanced Design System for MMIC design and PCB design. Previous designs completed are quad-band GSM power amplifier modules, 2.4-2.5GHz/5-6GHz wireless LAN power amplifiers, broadband power amplifiers, switch-filter designs, RFID detectors, RF/DC converters and filter designs.

Front End Module Design

We use Agilent's Advanced Design System and Orcad PCB Editor for Front-End Module designs. Previous design examples are high-volume switch filter modules, dual-band 802.11a/b/g front end modules, 76GHz car radar module, RFID card design.

RFID Detector System

On an NSF Phase I & II grant to one of our clients, operating as a subcontractor, the RF Diagnostics design team created a complete RFID MEMS resonator detection system. Both hardware and software were developed under this

LEARN MORE   contract. The work was published at an SPIE conference in July 2010.

NISKAYUNA, NY 12309 USA

518-810-6560

INFO@RFDIAGNOSTICS.COM

TOP

Sebs World

- Home
- DA14 Asteriod
- Secret space plane
- Solar storm
- End of the world?
- UK monitoring
- Tracking
- Save money
- Get out of the EU
- UFO's
- Magnetic energy
- RF energy
- Radiant energy
- Perpetual
- Easy Steam!
- Irony- defined
- Funny Pics
- Internet Problems
- Star Wars
- Chat room
- Live music
- Kevzone
- Facebook Cheats
- Crashes
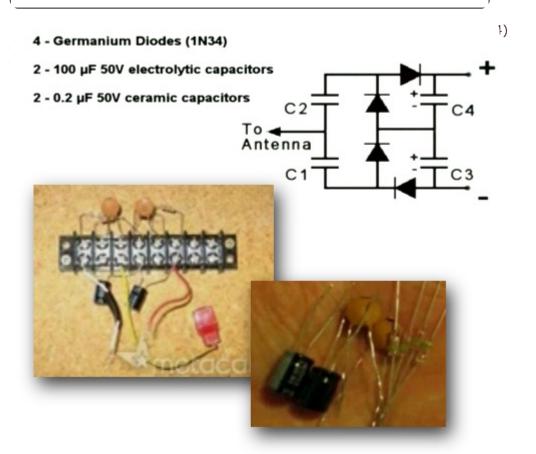
About us
Private Policy
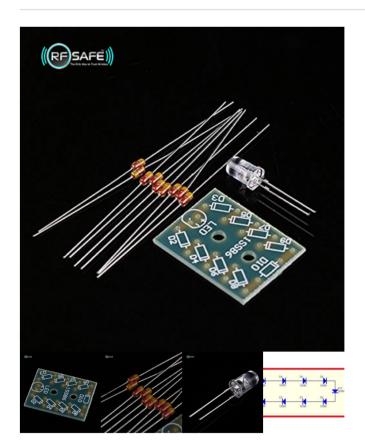
RF energy..

This can extract some electricity from the air!

FREE ENERGY FROM AIR CIRCUIT PART 8/24

4 - Germanium Diodes (1N34)

2 - 100 μF 50V electrolytic capacitors

2 - 0.2 μF 50V ceramic capacitors

DIY RF METER: LED IS 100% POWERED BY RF RADIATION USING DETECTOR DIODES

$8.99

| 1 | Add to cart |

SAFE KIDS: LED Cell Phone Radiation Detector For DIY Science Project & Experiments || Test Phone Shields

NO batteries. it uses ONLY RF radiation power from cell phone. Yes, see with your own eyes how powerful cell phone radiation is when it powers a light from thin air. RF Safe's DIY circuit is the simplest detector for microwave radiation. The RF diodes convert microwaves from a cell phone to DC current to power an LED which emits visible light warning you of possible danger.

SKU: 636684233801 Categories: Cell Phone Radiation Safety, D-I-Y Tags: 1SS86 detector diode, Cell Phone Radiation Meter, germanium diode, LED RF Meter

A concurrent 915/2440 MHz RF energy harvester

L Fadel, L Oyhenart, R Bergès, V Vigneras, T Taris

HAL Id: hal-01474918

https://hal.archives-ouvertes.fr/hal-01474918

Submitted on 7 Mar 2017

A concurrent 915 MHz / 2440 MHz RF Energy Harvester

L. Fadel[1], L. Oyhenart[1], R. Bergès[1], V. Vigneras[1], T.Taris[1]

[1] Laboratoire IMS, UMR - 5218, Université de Bordeaux, 33405 Talence Cedex, France.

This paper presents the development of two dual-band RF harvesters optimized to convert far-field RF energy to DC voltage at very low received power. The first one is based on a patch antenna and the second on a dipole antenna. They are both implemented on a standard FR4 substrate with commercially off-the-shelf (COTS) devices. The two RF harvesters provide a rectified voltage of 1V for a combined power respectively of -19.5 dBm at 915 MHz and -25 dBm at 2.44 GHz and of -20 dBm at 915 MHz and -15 dBm at 2.44 GHz. The remote powering of a clock consuming 1V/5µA is demonstrated, and the rectenna yields a power efficiency of 12 %.

I.   INTRODUCTION

Today's the society is evolving toward creating smart environments where a multitude of sensors and devices are interacting to deliver an abundance of useful information. Essential to the implementation of this Internet Of Things (IOT) is the design of energy efficient solutions aiming toward a low-carbon-emission, namely green, society. Within this context, the energy harvesting appears as an alternative to provide systems with self-sustained operation. Many electronic devices operate in conditions where it is costly, inconvenient, or impossible to replace the battery. Examples include sensors for health monitoring of patients [1],[2], aircraft or building structural monitoring [3],[4], sensors in natural, industrial or hazardous environments, etc. The scavenging of natural ambient energy requires some specific conditions such as: daylight for solar energy [5], breeze for wind energy or motion for kinetic energy [6] to name a few. As consequences the exploitation of natural source does not fit with many cases of applications. On the other hand the electromagnetic (EM) [7], or Radio-Frequency (RF), energy is a human made source that is not dependent of weather conditions nor the daytime. It is so very attractive for wireless powering of remote devices. Furthermore the ever growing of commercial and personal wireless installations opens up to a 24 hour a day available energy in the vicinity of any human activity areas. The schematic of a general Wireless RF Power Transmission (WPT) system is shown in Fig.1. We talk here about far-field RF energy transmission [8], which is different from near-field RF energy transmission [9]. This later including inductive, capacitive or resonant coupling is a close contact transmission and is not relevant for remote devices. In Fig. 1, the receiver antenna

collects the EM energy radiated by a RF source, and converts it into a RF signal. This RF signal is transferred to the rectifier by an impedance matching network, to be converted into DC power, which is further accumulated in a storage element. The main purpose in the deployment of WPT systems is the development of compact and efficient solutions. Most of the challenge concerns the implementation of harvesting modules, especially the antenna as its design defines the scavenging capability and the size of the RF harvester. At low frequency the transfer of energy is efficient, but the antenna footprint is large. To address the trade off between the efficiency of the WPT and the size of the modules, the frequency band located in the 433 MHz to 6 GHz frequency spectrums are preferred.
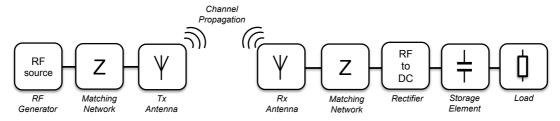
Fig. 1: Schematic of a Wireless RF Power Transmission (WPT) system

Over the last decade the research effort has focused on the development of WPT systems according two scenarios: the RF energy scavenging [10] and the RF energy transfer [11]. The two RF energy scavenging is an opportunistic collection of the RF ambient energy from the surrounding communication traffic. To improve the harvesting capability the scavenging RF harvesters are of wide-band type [12] and cover popular standards such as: DTV (470-610 MHz), GSM900, GSM1800, 3G (2.1GHz) and WiFi (2.4 GHz). Unfortunately these standards dedicated to convey wireless communications do not radiate a large RF power. As consequences the collected energy is weak, unpredictable and out of control. The RF energy scavenging remains a promising solution in the future as the increase of communication traffic could make it more reliable, and consistent with IOT applications. The second concept, namely RF energy transfer, assumes an identified source that is dedicated to perform the WPT. The amount of transmitted power is controlled by the source and the collected energy is larger than in scavenging approach. The licence-free Industry-Science-Medical (ISM) frequency bands located at 0.9, 2.4 and 5.8 GHz are usually exploited to support such a WPT scenario. Today the RF energy transfer in ISM Bands is not only promising, it becomes a reality as some pioneer companies propose some full kits: Powercast Corporation, AnSem and MicroChip to name a few. However there is still a lot of work to make the RF energy transfer an appropriate, low cost and easy-to-use solution for remote powering. One of the most critical point concerns the harvesting capability of the RF modules. So far the commercial kits referenced above only explore the 900 MHz ISM allocations to perform the WPT. This work proposes to demonstrate the interest of a concurrent harvesting at 915 MHz and 2.44 GHz. The design and implementation of a modified 4-stage doubler RF to DC converter, including a concurrent matching network, is first presented. The section III details the design of two types of multi-band antenna. The comparison between a single frequency and a multi-band WPT is exposed and the demonstration of the remote powering of a clock is reported as a case of application. To conclude a comparison of our results with the state of the art is exposed.

II. CONCURRENT RF TO DC CONVERTER

The Radio-Frequency IDentification (RFID) applications are the most popular systems exploiting the principle of RF energy transport. In passive RFID applications the reader transmits the RF power to the tag, and also sets up the communication. The RF to DC converter is designed to yield a maximum of power efficiency to the tag. Most of the time the reader and the tag are in line of sight and close to each other, these conditions improve the transmission of RF energy, the amount of power available at the tag antenna is large, typically between -15 dBm and -20 dBm. In RF energy harvesting the scenario is different. The distance between the RF source and the RF harvester ranges from 0.5 meter to 10 meters. The amount of collectable power is low, from -10 dBm to -25 dBm, and the remote powering is difficult. The RF harvesters are supposed to collect and to store the energy during a long period of time. Once the level of stored energy is large enough, it can be released to the application. For these reasons a rectifier dedicated to RF energy harvesting is first designed to yield a maximum of sensitive to increase its scavenging time and capability.

A) Rectifier Architecture

The rectifier architecture is based on voltage multipliers to provide an adequate output DC voltage. The architecture of the RF to DC converter, reported in Fig.2, includes a matching network based on a L-section, and a N-stage voltage multiplier based on Schottky diodes from Avago (HSMS285). The choice of the Schottky diode is very important in the design of the rectifier. A key parameter is its threshold voltage V_{TH}. When only low power levels are available in the environment, the amplitude of the incident signal may be close to or even below this voltage. Below this voltage value, the diode will no longer conduct and the losses become predominant. For COTS devices the two Schottky diodes performing the best conversion efficiency in a 2.4 GHz range are HSMS-2850 from Avago and SMS-7630 from Skyworks [13].
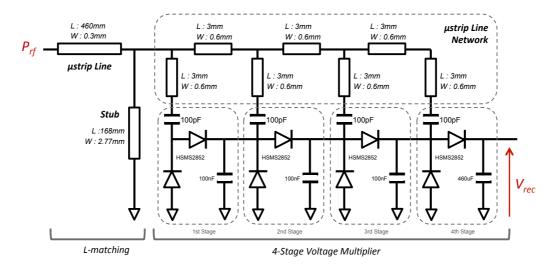
Fig. 2: Architecture of the RF to DC converter

Focusing on the sensitivity, the RF to DC converter is designed to maximize the rectified voltage for an input power close to -20 dBm. The optimum number of stage is fixed to four according [14]. The footprint of each voltage doubler imposes the micro-strip line network. The micro-strip lines namely "junction" is set to minimum length, the micro-strip lines "access" are used as an additional degree to tune the input-matching network. Indeed, the L section in combination with the micro-strip distributed network is equivalent to a T section (Fig.3). Many combinations of Z_1, Z_2, Z_3 can achieve input matching at 900 MHz or 2.4 GHz. Some of them are very close for each frequency, so we choose one that allows a return loss (< - 10 dB at least) both at 900 MHz and 2.4 GHz.
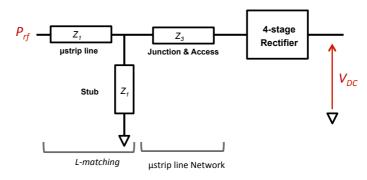
Fig. 3: Topology of the input-matching network

The equivalent narrow band model of the matching network is proposed for each frequency (Fig.4). At 915 MHz, the voltage multiplier, including the rectification stages and the micro-strip line network, is modelled with a shunt capacitor (5pF) and a shunt resistor of 270 Ω (Fig.4a). The stub, (Fig.3) is equivalent to an inductor (Fig.4a), which compensates the shunt capacitor. The input micro-strip line, (Fig.3), is a quarter wave impedance transformer, (Fig.4a) it converts the 270 Ω into 50 Ω. At 2.44 GHz the micro-strip line network distributing the RF signal to the voltage doublers (Fig.3), becomes inductive (Fig.4b). The stub is equivalent to a shunt capacitor of 120fF, its effect is negligible. The impedance transformation is actually performed by the input micro-strip line, which is modelled by a shunt capacitor (0,6 pF) and a series inductor of 5.6 nH.
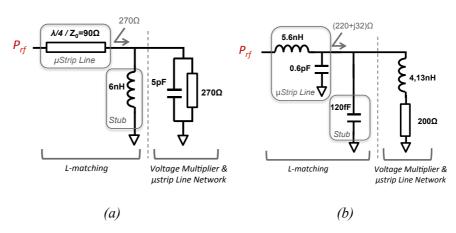
Fig. 4: Equivalent model of the input matching network at 915 MHz (a) at 2,44 GHz (b)

To study the impact of the power on the diode, and input matching, behaviour the return loss of the 4-stage rectifier has been measured and plotted (Fig.5) for various input power P_{rf} at 9000MHz.
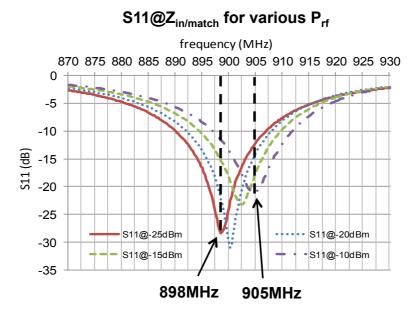
S11@$Z_{in/match}$ for various P_{rf}

Fig. 5: Measured S_{11} of a 4-stage voltage multiplier with a L section at 900 MHz for various input power P_{rf}

As illustrated in the figure 5, the input return loss is not strongly affected by the input power if P_{rf} < -15 dBm. The RF harvesters developed in this work are dedicated to collect power from -15 to -25 dBm. Over this range the diode model can be considered as stable, and the slight frequency shift is still covered by the antenna bandwidth.

B) Rectifier Characterization

The power efficiency and the power sensitivity are two conversion characteristics of importance in RF harvesters. However, the RF harvester operating at low power level accumulates the energy in a storage element, to further release it to the application. In such accumulation mode the power sensitivity becomes more important than the power efficiency.

For the characterization the rectifier is not connected to a load. The load represents the equivalent impedance of the application (clock, sensor) to power. The effectiveness of RF-DC conversion of the rectenna and its DC ouput voltage varies depending on the load value. The rectifier is first characterized in a single tone mode, 915 MHz and 2.44 GHz respectively, and then in a dual-band mode. Measurements of the unloaded rectified voltage versus various input power P_{rf} are reported in Figure 6.
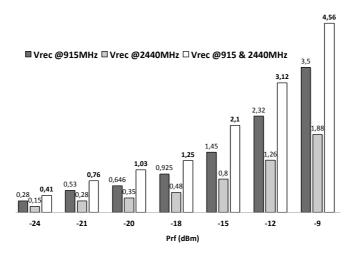
Fig. 6: Unloaded rectified voltage for various input power

To compare the results of the two considered tones, the target is fixed to a value of 1V. In a single tone mode, the required P_{rf} to rectify 1V is close to -18 dBm at 915 MHz, and would be larger than -15 dBm at 2.44 GHz. In a dual-band mode the circuit only needs a power P_{rf} of -20 dBm at each frequency. The dual-band rectification significantly improves the power sensitivity. The reverse breakdown voltage of the HSMS285 Schottky diode limits the input power to - 9 dBm, for which V_{rec} is 4,56 V.

III. ANTENNA DESIGN

To meet the low-cost constraints, the RF energy harvester will be implemented on a single low cost substrate, an FR4 PCB. For the antenna, there are more efficient substrates with a higher permittivity to reduce the size of the antenna or a lower losses but their cost is much higher than the improved performance. These powerful substrates fail to build low cost energy harvesters. This section exposes the design of two dual-band antennas implemented on a 1.6 mm FR4 printed circuit board. The fabrication uses a mechanical etching process with a 200µm resolution. We have chosen two complementary antenna topologies with a directional and omnidirectional radiation pattern.

A)   Dual - band patch antenna

Emitting and receiving antennas do not usually meet the same constraints. Mobile devices such as smartphones and tablets use compact antennas (ifa, pifa…) to address the trade off between performance and size. Base stations can afford large efficient radiating elements (omnidirectional or directional antennas depending on the application). For energy harvesting purpose, micro-strip patch antennas are commonly used [15-17]. A rectangular micro-strip patch antenna (RMPA) is first developed to suit with both low cost technology of implementation and co-integration with the rectifier. Based on the cavity-model approximation, the resonant frequencies of the RMPA for the TM$_{mn}$ mode is described in (1).

$$f_{mn} = \frac{c}{2\sqrt{\varepsilon_r}} \sqrt{\left(\frac{m}{L}\right)^2 + \left(\frac{n}{W}\right)^2}$$

(1)

W, L are the patch dimensions, c = 3.10^8m/s

The antenna dimensions, 68 cm^2 (8.8 x 7.8 cm), described Fig.7a are dependent to the frequency bands and the feed location is selected to only excite the fundamental modes TM_{01} and TM_{10}. Those modes permit to obtain a large aspect ratio (W/L = 2,7) but reduce the performance of the RMPA. On the other hand, TM_{01} and TM_{30} modes require an aspect ratio close to one but offer beneficial radiation patterns for our application. The RMPA is fed by a probe whose position (x,y) adjusts the matching both at 915 MHz and 2.44 GHz. This two operating bands of the proposed antenna are on cross polarization planes. The geometric parameters of RMPA have been optimized with an approximate model, the TL model [18], and with a full wave method. Details of the two approaches have been studied in [19]. The return loss of the RMPA is better at 915 MHz than 2.44 GHz because the maximum impedance of TM30 mode is 31 Ω [19]. The TM30 mode does not achieve 50 Ω because it is not a fundamental mode. This antenna has a maximum gain of 1.3 dB at 915 MHz (Fig.7b) and 2.5 dB at 2.44 GHz (Fig.7c). This two operating bands of the proposed antenna are on cross polarization planes.

The realized gains of the dual band patch antenna are lower than the classical patch antenna because the radiating efficiency is low, 60% for the TM_{01} mode and 30% for the TM_{30} mode. The FR4 substrate has a loss tangent of 0.02. The radiation efficiency of the dual band patch antenna is highly dependent of the substrate losses.
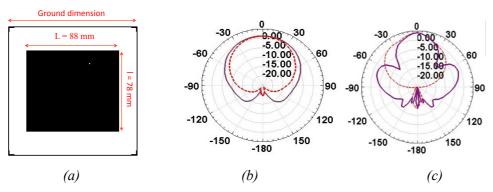
Fig. 7. Layout of the RMPA antenna (a) - Radiation pattern at 915MHz (b) and 2.44 GHz (c) Solid and dashed lines correspond to E-plane and H-plane respectively

B) Multi-band arm dipole antenna

The second antenna is a multi-band dipole type composed of three arms. Its dimension is about 23 cm^2 (11.1x2.1 cm), Fig.8a. Each arm is designed to work at one band of frequency. The longer one is for the 915 MHz, the middle one, not useful in our case, is for the 1.4 GHz and the last one, the smaller, is dedicated to 2.4 GHz [20].

All the geometric parameters have been optimized with a full-wave method in order to be matched both at 915/ 2440MHz. On Fig.6b and 6c, the radiation pattern is plotted for the elevation plane (orthogonal to substrate) of the simulated antenna. The maximum gain is 0.5 dB at 915 MHz and 3.4 dB at 2.44 GHz at 90°, on the substrate plane. The radiation efficiency is 99% at 915 MHz and 95% at 2.44 GHz.
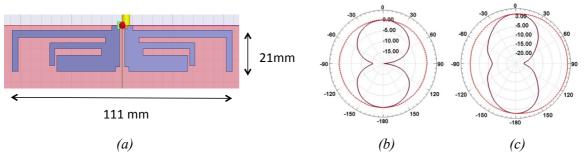
Fig. 8. Layout of the multi-band arms dipole antenna (a) - Radiation pattern at 915 MHz (b) and 2.44 GHz (c). Solid and dashed lines correspond to E-plane and H-plane respectively

It is interesting to compare the characteristics and performances of the two types of antennas. Although the radiation efficiency of the dipole antenna is better than the patch antenna, the antenna gains are similar because the high directivity of the RMPA antenna compensates the low values of radiation efficiency. When there are no cost constraints, it is interesting to use high performance substrates for the design of RMPA antennas because they improve the radiation efficiency and consequently antenna gain.

 Moreover, the integration of the antenna with the rectifier will not be made in the same way. Considering the patch antenna, the rectifier can be integrated on the ground plane allowing a more compact solution. The dipole antenna, which is ground plane free, is less sensitive to the surrounding environment in our case. The performance of the dipole antenna and especially the radiation efficiency are very weakly dependent of the substrate characteristics. The design of a dipole antenna can be easily reuse with other material such as Kapton®, paper, Plexiglas to name a few.

IV. WIRELESS POWER TRANSMISSION

This part presents the measurement results of the assembled RF harvesters in the context of Wireless Power transfer. The two dual-band harvesters are realized with COTS devices such as HSMS diodes and capacitors. Those elements are reported by heat-treating. The RF to DC converter board, including the matching network and the rectifier, is reported on the backside and connected to the radiation part, on the front side, through a via (Fig.9a). The dipole antenna is connected to the rectifier circuit using SMA connector (Fig.9b).
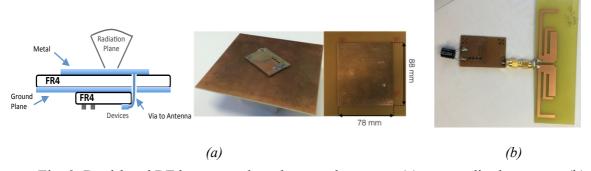
Fig. 9. Dual-band RF harvesters based on patch antenna (a) – arms dipole antenna (b)

For the dual-band RF harvester based on patch antenna, the return loss, S_{11}, is measured for an input power of -20 dBm with a HP8720 network analyser. The patch antenna, the rectifier and the dipole antenna are centered at 915 MHz and 2.44 GHz with a low return loss ($S_{11}< -15$ dB), Fig.10. The return loss of the RMPA antenna is better at 915 MHz than 2.44 GHz because the maximum impedance of TM30 mode is 31 Ω (Fig. 6 and Fig. 7 of [16]). The TM30 mode does not achieve 50 Ω because it is not a fundamental mode.
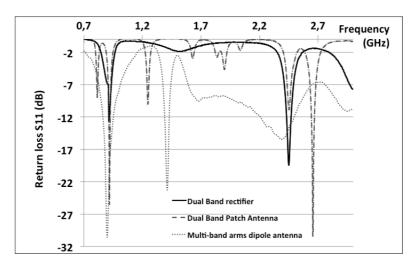
Fig. 10. Measured return loss S_{11} of the dual-band rectfier, patch and arms dipole antenna

A)   Remote Powering and Power Efficiency

The rectenna is connected to a clock, which mimics a low power application. The remote powering of this clock is performed in a furnished room of the lab according the schematic of Fig.11. The distance between the source and the antenna is fixed to 2 m. The clock is turned on for different scenarios of transmitted power. For each combination of power proposed in Fig.12, the RF power is first measured with a calibrated antenna and a power meter. Then, the rectenna is measured and P_{eff} is the ration between the power delivered to the load (here the clock) and the power available at the antenna.
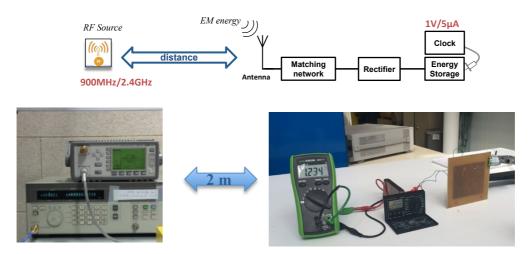
Fig. 11. Schematic and picture of the scene of remote powering of a clock

The power efficiency of the patch and the dipole rectenna is worked out from these experiments and reported in Fig.12. The power efficiency η is defined as the ratio between the DC power delivered to the clock and the RF power collected by the antenna.
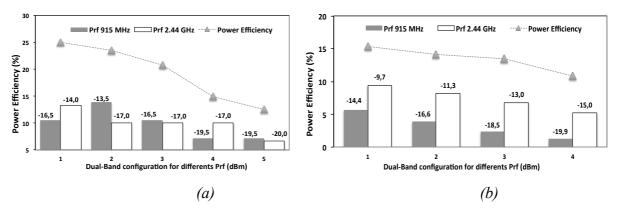
Fig. 12. Power efficiency of the dual-band RF harvester based on the patch (a) and the arms dipole (b) antenna

The minimum power required to turn on the clock with the patch-based harvester, Fig. 10.a, is a two tone signal featuring: -19.5 dBm at 915 MHz and -20 dBm at 2.44 GHz. At this point, the power efficiency is 12.5 %, which corresponds to a DC output power of 2.7 µW/1V. A maximum efficiency of 24 % occurs for a combined power of -16.5 dBm at 915 MHz and -14 dBm at 2.44 GHz. The harvester is able to deliver a DC ouput power of 15 µW. The harvester based on the arm dipole antenna needs a minimum power of -19.9 dBm at 915 MHz and -15 dBm at 2.44 GHz at the antenna to turn on the clock. For these conditions of remote powering, the efficiency of the harvester is 11 %. It delivers a DC power of 3.8 µW/1.15 V. The maximum power efficiency, 15.5 %, yields for an input power of -14.4 dBm at 915 MHz and -9.7 dBm at 2.44 GHz, the DC output power is 21 µW.

This scenario of remote powering figures out that the harvester based on the patch antenna exhibits a better power efficiency than the harvester combined with the dipole element. This difference is due to the antenna gains. Referring to Fig. 7 and Fig. 8, the gain of the patch antenna is larger (+0.8dB) at 915 MHz and lower (-0.9 dB) at 2.44 GHz than the dipole element. However the rectifier, referenced in [16], achieves a power efficiency of 17% at 915 MHz and only 5% at 2.44 GHz for an input signal of -15 dBm. As consequences the patch-based harvester is able to extract more power from a 915 MHz signal than the dipole-based harvester can do at 2.44 GHz. For this reason the overall efficiency of the patch harvester is better.

B)   Power Sensitivity

The power sensitivity is measured with the same scenario of Fig.11 but the clock is disconnected. The output voltage is reported for different combination of collectable power at the antenna in a dual-band configuration.

(a)   *(b)*
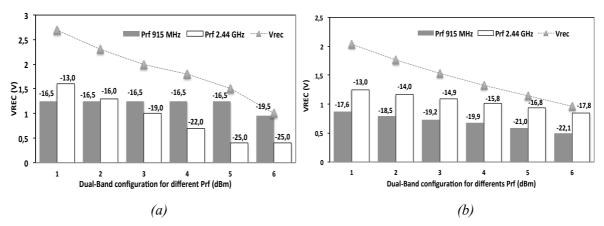
Fig. 13. Rectified voltage of the RF harvester based on patch (a) and arm dipole (b) antenna

To rectify a 1V DC voltage, the patch-based harvester, Fig.13a, requires a two dual-band configuration: -19.5 dBm at 915 MHz and -25 dBm at 2.44 GHz, which is equivalent to an input power of -18.4 dBm (or 14 μW). For the same purpose the dipole-based harvester, Fig.13b, needs a dual-tone of -22.1 dBm at 915 MHz and -17.8 dBm at 2.44 GHz. The equivalent input power of this 2-tone signal is -16.5 dBm (or 22 μW). The patch-based harvester exhibits a better sensitivity than the dipole harvester for the same reason exposed in the part A of this section. In Fig.6, which reports the power sensitivity of the rectifier part only, the overall sensitivity is almost the same for the dipole harvester. It is improved by 1.8 dB for the patch harvester due to the additional gain of the antenna at 915 MHz.

C)   Discussion and Comparison with the state of the art

An important characteristic of a remote powered device is its size. Indeed it is expected to be as small as possible to make it unobtrusive to our closest environment. In a scenario of RF harvesting the antenna footprint determines the compactness of a harvester operating Ultra High Frequency (UHF) bands.   To complete the comparison between the two harvesting modules developed in this work, two figures of merit, FOM_{sens} and FOM_{eff}, including the size of the antenna, are proposed in (2) and (3).

$$FOM_{sens} = \frac{V_{REC@Psens}(V)}{\frac{P_{sens}(\mu W)}{100 \mu W} \cdot \frac{A_{ant}(cm^2)}{100 cm^2}} \qquad (2)$$

With: P_{sens} the input RF power required to provide $V_{REC@Psens}$ the unloaded rectified output DC voltage, and A_{ant} the area of the antenna.

$$FOM_{eff} = \frac{\eta(\%)}{\frac{P_{eff}(\mu W)}{100 \mu W} \cdot \frac{A_{ant}(cm^2)}{100 cm^2}} \qquad (3)$$

With: P_{eff} the input RF power required to achieve η the overall power efficiency.

FOM_{sens} and FOM_{eff} do not represent the same scenario of application. The FOM_{sens} illustrates the capability of the rectenna to start collecting energy and store it in an element such as a capacitor or a battery to further release it. FOM_{eff} demonstrates the capability of the RF harvester to yield "on time powering": the rectenna is connected to an application and

supply it on time. Both are reported in the Table I, which also includes some references of the state of the art. The ability of the proposed rectenna to simultaneously operate in two frequency bands, significantly improves the power sensitivity.

Table I. Comparison with the state of the art

Ref.	Freq (GHz)	Efficiency (%@P_{rf})	Sensitivity (V_{rec}@Prf)	Number of stage	Schottky diodes	Size (cm^2)	FOM Sens	FOM Eff
[21]	0.9	15% @ -10dBm	0.75V @ -10dBm	1	SMS-7630	15×15	1.9	6.6
[21]	2.4	9% @ -13dBm	0.9V @ -13dBm	1	SMS-7630	15×15	1.25	8
[22]	2.45	10.5% @ -20dBm	0.075V @ -20dBm	1	SMS-7630	3.4×3.4	6.5	905
[23]	0.915/ 2.45	14%@ -20/-20	0,36V@ -10/-10	1	SMS-7630	6×6	0.5	185
[24]	1.8/2.2 /2.5	55%@ -10dBm	300mV@ -32dBm 3tones	1	SMS-7630	7×7	20	112
This work Dipole	0.915/ 2.44	**11% @ -20/-15dBm**	**1V @ -22/-18dBm**	4	HSMS-2850	2.1×11	**19.8**	136
This work Patch	0.915/ 2.44	**12.5% @ - 19.5 dBm/ -20dBm**	**1V @ -19.5dBm/ -25dBm**	4	HSMS-2850	7.8×8.8	**10.5**	100

According the Table I, the patch-based harvester exhibits the highest sensitivity to rectify 1V with a dual-tone featuring: -19.5 dBm at 915 MHz and only -25 dBm at 2440 MHz. The FOM$_{sens}$ represents the trade-off between the sensitivity performances of a rectenna and the antenna area. The FOM$_{eff}$ rates the efficiency performances to the antenna area. For these two figures of merit, the rectenna based on the multi-arm dipole element yields the best trade-off, both for FOM$_{sens}$ and FOM$_{eff}$, compared to the patch-based solution. This dual tone and multi-arm dipole harvester is close to the work proposed in [24] which exhibits the highest FOM$_{sens}$ reported so far in the literature to our knowledge.

V. CONCLUSION

The range of power collectable in a scenario of RF harvesting varies from -15dBm to -25dBm. To address this purpose the rectenna proposed in this work are optimized to operate at a RF input power close to -20 dBm (or 10 µW). To further improve the ability to collect the RF energy, these rectenna, developed with Schottky diodes HSMS285 from Avago, perform a concurrent harvesting in the 915 MHz and 2.44 GHz ISM bands. The harvester including a patch antenna implemented on a 1.6mm FR4 PCB achieves the highest sensitivity. It provides a 1V-rectified voltage for a dual-tone excitation of -19.5 dBm at 915 MHz and -25 dBm at 2.44 GHz. For these conditions of operation the rectenna yields a power efficiency of 12.5%. To take into account the dimensions of the haverster, two figures of merit, FOM$_{sens}$ and FOM$_{eff}$ including the size of the antenna, respectively related to the power sensitivity and the power efficiency are proposed. The rectenna developed with the arm dipole element exhibits the highest figures of merit. A case of application is proposed with the remote powering of a

digital clock consuming 1V/5µA. The patch based harvester turns on the device with a dual tone excitation at the antenna of -19.5 dBm at 915 MHz and -20 dBm at 2.44 GHz. For the same scenario the harvester connected to the multi-arm dipole element needs a power of -22.1 dBm at 915 MHz and -17.8 dBm at 2.44 GHz.

REFERENCES

[1] T. Paing, A. Dolgov, J. Shin, J. Morroni, J. Brannan, R. Zane, and Z. Popovic, "Wirelessly powered wireless sensor platform'' in Eur. Microw. Conf. Dig, Munich, Germany, Oct. 2007, pp. 241-244.

[2] J. Bernhard, K. Hietpas, E. George, D. Kuchima, and H. Reis, ''An interdisciplinary effort to develop a wireless embedded sensor system to monitor and assess corrosion in the tendons of pre-stressed concrete girders,'' in Proc. IEEE Top. Conf. Wireless Commun, 2003, pp. 241-243.

[3] C. Walsh, S. Rondineau, M. Jankovic, G. Zhao, and Z. Popovic, ''A conformal 10-GHz rectenna for wireless powering of piezoelectric sensor electronics,'' in IEEE MTT-S Int. Microw. Symp. Dig, Jun. 2005, pp. 143-146.

[4] X. Zhao, T. Qian, G. Mei, C. Kwan, R. Zane, C. Walsh, T. Paing, and Z. Popovic, ''Active health monitoring of an aircraft wing with an embedded piezoelectric sensor/actuator network,'' Smart Mater. Struct, vol. 16, no. 2007, Jun. 2007, pp. 1218-1225.

[5] K. Lin, J. Yu, J. Hsu, S. Zahedi, D. Lee, J. Friedman, A. Kansal, V. Raghunathan, and M. Srivastava, "Heliomote: Enabling long-lived sensor networks through solar energy harvesting," in 3rd Int. Conf. Embedded Networked Sensor Syst., Nov. 2005, p. 309.

[6] J.A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," IEEE Pervasive Computing, Vol. 4, and Issue: 1, Jan-March 2005, pp. 18- 27.

[7] K. R. Foster, "A world awash with wireless devices", IEEE Microwave magazine, vol. 14, Issue 2, March 2013, pp. 73-84.

[8] T. Le, K. Mayaram, and T. Fiez, "Efficient far-field radio frequency energy harvesting for passively powered sensor networks," IEEE J. Solid-State Circuits, vol. 43, no. 5, May 2008, pp. 1287-1302.

[9] J. Hirai, T.W. Kim, and A. Kawamura, ''Wireless transmission of power and information for cableless linear motor drive,'' IEEE Trans. Power Electron, vol. 15, 2000, pp. 21-27.

[10] H. Nishimoto, Y. Kawahara, and T. Asami, "Prototype implementation of ambient RF energy harvesting wireless sensor networks," IEEE Sens. J., Nov. 2010, pp. 1282-1287.

[11] BN. Shinohara, "Power without wires," IEEE Microw. Mag, vol. 12, no. 7, Dec. 2011, pp. S64-S73.

[12] M. Pinuela, P.D. Mitcheson, S. Lucyszyn, "Ambient RF Energy Harvesting in Urban and Semi-Urban Environments", IEEE Trans. On Microwave Theory and Techniques, vol. 61, N°7, July 2013.

[13] Simon Hemour, Ke Wu "Radio-Frequency Rectifier for Electromagnetic Energy Harvesting : Developement Path and Future Outlook", Proceeding of the IEEE, Vol.102, No.11, November 2014.

[14] T. Taris, L.Fadel, L.Oyhenart, V. Vigneras, "COTS-Based Modules for Far-Field Radio Frequency ", IEEE NEWCAS, Paris, France, June 2013, pp. 1-4.

[15] Sika Shrestha, Sun-Kuk Noh, and Dong-You Choi, « Comparative Study of Antenna Designs for RF Energy Harvesting » International Journal of Antennas and Propagation Volume 2013, 10 pages.

[16] Z. W. Sim, R. Shuttleworth, M. J. Alexander, B. D. Grieve, « Compact patch antenna design for outdoor RF energy harvesting in wireless sensor networks, Progress In Electromagnetics Research, Vol. 105, 273–294, 2010.

[17] Naimul Hasan, Santu Kumar Giri, « Design of low power RF to DC generator for energy harvesting application », International Journal of Applied Sciences and Engineering Research, Vol 1, Issue 4, 2012.

[18] H. J. Visser, "Approximate Antenna Analysis for CAD", Wiley, 2009.

[19] R. Berges, L.Fadel, L. Oyhenart, V. Vigneras, T. Taris, "A dual Band 915MHz/2.44GHz RF Energy Harvester", EUMW, Paris France, Sept 2015.

[20] Y-Y.Lu, J. Guo, H-C. Huang, "Design of Triple Symmetric Arms Dipole Antenna for 900/1800/2450MHz Applications", Conf on Intelligent Information Hiding and Multimedia Signal Processing", 2014.

[21] D. Masotti, A. Costanzo, M. Del Prete, V. Rizzoli, "Genetic-based design of a tetra-band high-efficiency radio-frequency energy harvesting system," IET Microwaves, Antennas Propagation, vol.7, no. 15, June 2013, pp. 1254-1263.

[22] G. Vera, A. Georgiadis, A. Collado, and S. Via, "Design of a 2.45GHz rectenna for electromagnetic (EM) energy scavenging", Proc. IEEE Radio and Wireless Symp, 2010, pp. 61-64.

[23] K. Niotaki, S. Kim, S. Jeong, A. Collado, A. Georgiadis, and M. Tentzeris, "A compact dual-band rectenna using slot-loaded dual band folded dipole antenna," IEEE Antennas and Wireless Propagation. Lett. vol. 12, pp. 1634–1637, 2013.

[24] C. Song, Y. Huang, J. Zhou, J. Zhang, S. Yuan, and P. Carter, "A High-efficiency Broadband Rectenna for Ambient Wireless Energy Harvesting", IEEE Transactions on Antennas and Propagation, accepted May 2015.

Bibliographies

Ludivine FADEL received the Ph.D in electrical engineering from the University of Bordeaux, France, in 2004. She is an Associate Professor at Bordeaux University and her main research interests include wireless power transfer and energy harvesting, materials and printing technologies for RF application.

Laurent Oyhenart received the Ph.D. degree in electrical engineering from the University of Bordeaux, Bordeaux, France, in 2005. He is currently an Associate Professor with the University of Bordeaux and develops his research activities in the IMS laboratory. His research interests include computational electromagnetics, photonic crystals and antenna synthesis.

Romain Bergès is a PhD student at IMS Laboratory. He received his Master from Bordeaux University in Electronic, in 2014. His current research focuses on RF Energy Harvesting.

Valérie Vigneras is a full professor in the Polytechnic Institute of Bordeaux and conducts research in the IMS Laboratory on complex structures (mixtures, structured media, self-assembled media) both by simulation to predict or optimize their properties and by characterization before their integration in high frequency systems (antennas, electromagnetic compatibility, radar shieldness…).

Thierry Taris is a full professor at the Bordeaux Institute of Technolgy (Bx-INP). He joined the IMS lab in 2005 where is research interests are related to the Radio-frequency Integrated Circuits in Silicon technologies.

List of figures and tables

Fig. 1: Schematic of a Wireless RF Power Transmission (WPT) system.
Fig. 2: Architecture of the RF to DC converter.
Fig. 3: Topology of the input-matching network.
Fig. 4: Equivalent model of the RF to DC converter, 915 MHz (a) - 2,44 GHz (b).
Fig. 5: Measured S_{11} of a 4-stage voltage multiplier with a L section at 900 MHz for various input power P_{rf}.
Fig. 6: Unloaded rectified voltage for various inpout power.
Fig. 7. Layout of the RMPA antenna (a) - Radiation pattern at 915MHz (b) and 2.44 GHz (c) Solid and dashed lines correspond to E-plane and H-plane respectively.
Fig. 8. Layout of the multi-band arms dipole antenna (a) - Radiation pattern at 915 MHz (b) and 2.44 GHz (c). Solid and dashed lines correspond to E-plane and H-plane respectively.
Fig. 9. Dual-band RF harvesters based on patch antenna (a) – arms dipole antenna (b).
Fig. 10. Measured return loss S_{11} of the dual-band rectfier, patch and arms dipole antenna.
Fig. 11. Schematic and picture of the scene of remote powering of a clock.
Fig. 12. Power efficiency of the dual-band RF harvester based on the patch (a) and the arms dipole (b) antenna.
Fig. 13. Rectified voltage of the RF harvester based on patch (a) and arm dipole (b) antenna.
Table I. Comparison with the state of the art.

blog

Fun with Less Kilowatts: The Lectenna

By Vernon Trollinger, March 14, 2017, Energy Efficiency, Events & Fun, Family

Welcome to *Fun with Less Kilowatts*! We believe that science experiments at home can be a creative way to engage kids in learning while having fun. They can be educational AND great activities to keep your kids busy and away from the television. Each month, we'll feature a new science experiment that can be a great resource for parents and teachers.

The Lectenna

Did you know you that the energy from radio waves can be converted back into electric energy? It's true and you can make a *simple* two-component circuit that you and your kids can use to find 2.45 Ghz radio waves around your home. These are the same radio waves used by microwaves, WiFi network routers, bluetooth devices, smart phones, cordless phones, smart meters, and smart home systems that use Zigbee.

YouTuber pjaffeva posted this brilliant project showing how. Plus, there's also a link to a PDF set of instructions on Google Drive that you can download.

The Materials

- One HLMP-D150 Avago low current LED. Costs about 50¢/each. Available from: Amazon, Digikey, Mouser

- One Hitachi 1SS106 low capacitance Schottky diode. Pjaffeva recommends this Hitachi-made diode as being the only one in this shape that will work. But they're also hard to find. Currently, you can get them from LittleDiode's ebay store in the UK. Costs $6.38 but the bulk of that price is shipping from the UK.

You'll also need:

- An operating WiFi router
- A bamboo kabob skewer
- Two pieces of tape

The Directions

1) Look on the base of the LED and find the side with the flat edge. This marks the negative or cathode side. Carefully splay the two wire leads away from each other as far as they will go. Make sure the one adjacent to the negative side gets bent over the flat edge so identify it. You can also mark it with a little spot of colored nail polish.

Categories

Careers/Jobs
Charity & Community
Energy Efficiency
Events & Fun
Family
FAQs
Green
Home Improvement
Hurricane Prep
Moving
News
Recipes
Save Money
Small Biz
Special Offers

Subscribe via Email

[ ]
Subscribe

Subscribe via Email

[ ]
Subscribe

Subscribe

🔶 Subscribe in a reader

Archives

December 2017
November 2017
October 2017
September 2017
August 2017
July 2017
June 2017
May 2017
April 2017
March 2017
February 2017
January 2017
all archives

Featured Articles

Prepare for Christmas With Green Cleaning Methods
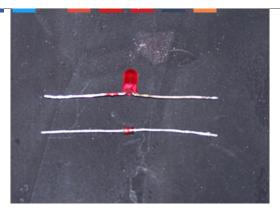
2) Next, find the band that runs around the Schottky diode. This identifies the negative end of this diode. The other end is the positive or anode end.

Using your fingers:

3) Twist the end of the negative or cathode wire from the LED onto to the positive or anode end of the Schottky diode.

4) Then twist the end of the negative or cathode end of the Schottky diode onto to the positive or anode end of the LED.

You have just made a "rectenna" and you'll need to keep the assembly and the wires straight as possible. The length of the wires should be about 6 cm long (a hair short of 2 3/8 inches).

5) Tape the two diodes to one end of the bamboo skewer.

6) With a WiFi router on and running, hold the diode near an antenna on the WiFi router.

The Result

The LED should light up as it draws closer to the antenna (signal). The further away it is, the weaker the signal and the less it can grab to light up.

The Science

The WiFi routers uses radio waves to transmit data to and from computers and other devices in your home. Radio waves behave like water in a glass, they ripple or oscillate at a certain number of times or frequency per second. A radio wave's wavelength is the distance covered by one complete cycle of a radio wave — from peak to valley.

Radio waves used by WiFi routers, bluetooth devices, microwaves, etc., broadcast signals in the 2.45 gigaherz (Ghz) range or 2.45 billion oscillations per second. That means the signal wavelength is about 12 cm. To pick up signal, however, you only need an antenna that's half as big - 6 cm. When the radio waves strike this antenna, called a 1/2 dipole ("di"=two), the energy excites the antenna's
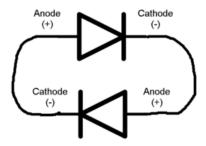
Follow Bounce: [f] [twitter] [in] [G+] [youtube] [pinterest] [t] [rss]   Share this on:

The "Rectenna": Because the radio waves cause voltage to flow one direction and then the opposite direction, the current is called an alternating current (AC). By connecting a diode to both poles of the antenna it converts the AC current to flow only in one direction, called direct current (DC). The antenna becomes a "rectenna".

How the diodes work: The diodes work as one-way valves for electricity, allowing voltage to flow forward but not back. In the case of the rectenna, AC current caused by the radio wave sends a positive charge down one of the antenna poles and then the other. By putting the Hitachi 1SS106 diode across both poles, the diode only lets the positive charge coming down one of the antenna poles through. The LED is actually a Light Emitting Diode and works the same way as a regular diode except that when it's on, it lights up. Meanwhile, Schottky diodes turn on at a lower voltage than other diodesand because the rectenna can only catch a fraction of the WiFi's broadcast power, the Schottky diode can rectify that little bit of AC to DC.

When we connected our diodes together, their polarity formed a loop that only allowed voltage to flow one way and thus light the LED. Brilliant job, pjaffeva!

Do you have any fun and kid-friendly science experiments you'd like to see us try for Fun with Less Kilowatts? Share with us in the comments!

Be Sociable, Share!

About Vernon Trollinger
A native of Wyomissing Hills, PA, Vernon Trollinger studied writing and film at the University of Iowa, later earning his MA in writing there as well. Following a decade of digging in CRM archaeology, he now writes about green energy technology, home energy efficiency, DIY projects, the natural gas industry, and the electrical grid.

Tags: family fun, Fun with Less Kilowatts, kids, science experiments

338K people like this. Sign Up to see what your friends like.

Follow @BounceEnergy   28.2K followers

Share 1   Tweet 0   Pin 0   G+

Follow Bounce: Share this on:

0 Comments   **Bounce Energy Blog**   1 **Login** ⌄

♡ **Recommend**   ⬆ **Share**   Sort by Best ⌄

👤 | Start the discussion… |

LOG IN WITH   **OR SIGN UP WITH DISQUS** ⑦

Ⓓ Ⓕ Ⓣ Ⓖ | Name |

Be the first to comment.

ALSO ON BOUNCE ENERGY BLOG

Choosing the Best Home Generator for Your Emergency Needs

1 comment • 3 years ago

T&B **T&B Electric LTD** — Nice article covering the difference between portable and permanent generators. Some homeowners "back feed" …

Take Care of Your Favorite Gadgets with Great Waterproof Tech

1 comment • 2 years ago

👤 **erna** — Vermon thanks for your post, great as always! Coolest is cool I got one of this :) what's funny I also have swimmo watch for swimmers …

Career Opportunity with Bounce Energy | PHP/Integration Developer

1 comment • 3 years ago

👤 **Chris Jones** — Sounds like a great opportunity.

5 Quick DIY Home Improvement Fixes to Get Your Home Through Winter

1 comment • 3 years ago

👤 **Steve Veach** — Excellent and informative blog post! I am a HVAC Contractor in Chicago and have found a new site to check in on. Thank …

✉ **Subscribe**   Ⓓ **Add Disqus to your site**Add Disqus**Add**   🔒 **Privacy**   **DISQUS**

Follow Bounce: f t in G+ ▶ P t 📶   Share this on:
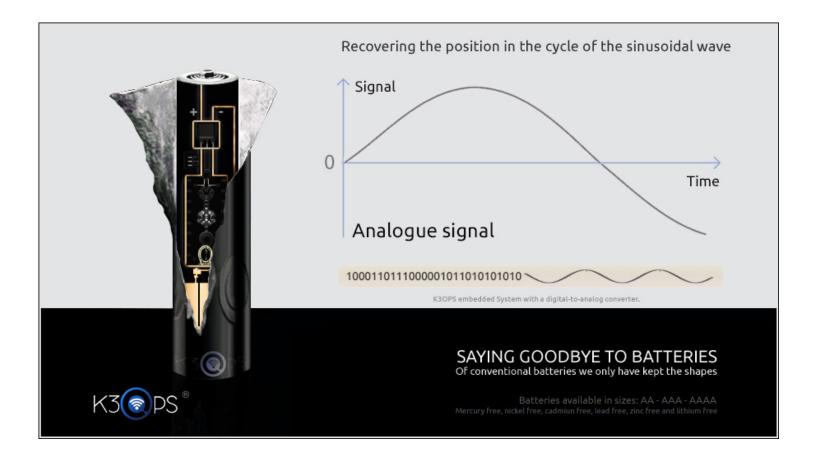
An electromagnetic field uses the photon as an elementary particle to transmit force. It combines:

- **A magnetic field** *force resulting from the movement of loads µT.*
- **An electric field** *force created by the attraction of repulsion loads, measured in volts per meter - V/m.*

With an energy determined according to the speed of light, the RF are by far the best medium to transmit any kind of information.

The multiplication of wireless communications systems in our environment ensures sufficient microwave leakages to harvest from the ambiant and enough energy to convert into DC electricity. Electromagnetic fields are everywhere and since they carry energy, they became the best candidate to deliver an endless source of renewable energy.

$\vec{B}$ is the magnetic induction expressed in T refered to **Nikola Tesla**, "Father of Free Energy", which is at the origin of the electromagnetism.

Using meta-materials combined with nanotechnology has deeply increased the performance and miniaturization of rectennas embedded in K3OPS system. Our products operate autonomously, offering an endless supply of green energy in a respectful and environment-friendly approach.

K3⦿PS®

| HOME |

principles and beautiful sentences. At that point in time I would have simply been Xin, a child like any other...

But one day I dreamed. I woke up far away, somewhere else, in another past. I decided to change my destiny and even if I was supposed to become a mathematician, I eventually decided to create, because already as a child my heart was chasing the stars.

Of the hundreds of directions shown to me after graduating, only one captured my attention: a single goal... Build the impossible for a safer world. So, over the past 3 years, our real challenge to overcome for all RF Energy Harvesting technics was to optimize electricity conversion. The massive proliferation of wireless telecommunication systems since the past two decades brought a saturation of the electromagnetic fields with a constant growth of 15% every year in our environments. As a result, this situation reversed the base problem that makes today Harvesting RF Energy a game changer. The key was the Power Management System.

We are far beyond the conversion constraint and performance by controlling "RF-interferences", by harvesting different frequencies *from near and far*, by using Metamaterials combined with nanotechnologies. We dramatically have improved power conversion efficiency and reduced the size of our Energy Harvesting systems embedded in all K3OPS' products.

Thanks to Nikola Tesla, my inspiring mentor, K3OPS' products have reached by far their objectives in terms of converting and performance, offering an endless efficient source of green energy, reliable in an environmentally friendly approach.

Xin WEI Co-Founder of K3OPS technology with Alexandre Despallieres

| Low Pass Filter | ⇨ | Matching network | ⇨ | Rectifier | ⇨ | DC Output |

The Rectenna was invented in 1964 by William C. Brown, patented in 1969. It is a rectifying antenna used to convert microwave energy into DC. A simple Rectenna consists of a dipole antenna with an RF diode connected across the dipole elements. The diode rectifies the AC current induced in the antenna to produce DC power.
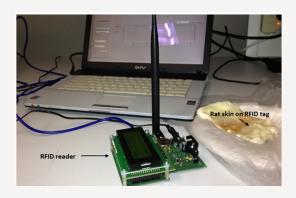
Centre for Collaboration in Electromagnetics and Antenna Engineering is a centre for research collaboration established to foster research in electromagnetics and engineering conducted by Macquarie University researchers and prominent collaborators from external institutions, including people from industry and overseas ins We have state of the art research ficilities including a NSI-700S-50 spherical nearfield anechoic chamber, 3 Vector Network Analyzers (capable up to 22GHz, 50GHz, and respectively), state-of-the-art embroidery machine DreamCreator-XE-VM5100 for embroided antennas, Dielectric Characterization Kit - High Temperature Probe (20 20GHz), as well as several in-house and commercial licenced softwares.

Our recent and ongoing research projects are:

- Wireless Implantable Bio-Telemetry System and Miniature Antenna Design
- Wireless Freedom for Lab Rats
- Flexible and Wearable Antennas for Biomedical and Healthcare Applications
- Characterizing Properties of Carbon nano Tubes at Microwave Frequencies
- Electromagnetic Band Gap (EBG) Resonator Antennas
- Leaky-Wave Antennas for Advanced Wireless Systems
- Shared Aperture Arrays for Space Borne Applications
- Novel Dielectric Resonator Antennas
- Super Wideband Antennas
- Focal Plane Arrays for Radio Astronomy
- Archimedian Spiral Metamaterials
- Frequency Selective Surfaces for Energy-Saving Glass Panels
- Dual-Band Artificial Magnetic Conductor (AMC) Surfaces
- Antenna Technologies for Ultrawideband (UWB) Systems
- High Gain Antennas with Planar Surface-Mounted Short Horns
- Photonic Crystal/EBG Based Horn Antennas
- Broadband Microstrip Patch Antennas for Wireless Computer Networks
- Theoretical Analysis of Photonic Crystal Structures
- New Closed-form Green's Functions for Microstrip Circuits and other Layered Structures
- Integrated-design of Hybrid-resonator Antennas for Broadband Wireless and other Communication Systems
- Singularity-enhanced Finite-different Time-domain (FDTD) Method for Diagonal Metal Edges, Strips and Films
- Low-profile Dielectric-resonator (DR) Antennas

Wireless Implantable Bio-Telemetry System and Miniature Antenna Design

The two major challenges associated with the conversion of a wireless system operating in air to an implantable version, antenna detuning and biocompatibility, are a in a coherent way. An RFID-based biomedical telemetry system designed for free-space operation was chosen as the starting reference. A new, pin-compatible, spa antenna with a ground plane was designed, fabricated and tested, to replace the original "free-space" antenna in the active RFID tag without making any other chang tag circuit, such that the tag would function well when it is placed under rat skin and fat. Biocompatibility and potential antenna detuning due to rat tissue variati addressed in the design process, without significantly increasing the tag physical height, by applying a thin coating of biocompatible material directly over the ante operation of the medical telemetry system was successfully demonstrated, with the tag placed under rat skin and fat, and its range of 60-72 cm was found to be suf support medical research experiments conducted with rats in cages. Due to the biocompatible coating over the antenna, antenna matching is very insensitive to ch tissue dielectric constants and thickness. The footprint of the new antenna is 33% less than that of the original antenna, its measured 10 dB return-loss bandwidth is 10 11%, and overall efficiency is 0.82% at 920 MHz.
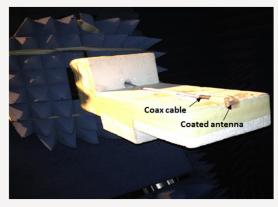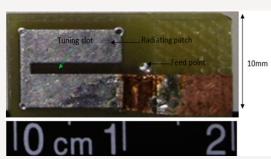
Wireless Freedom for Lab Rats

We are developing a fully implantable wireless telemetry system. This is a joint research project with BCS Innovations and the Australian School of Advanced Medicine (will be first used in the research conducted in ASAM, with rats, on hypertension. To date the major method of controlling hypertension is through the use o pharmaceuticals. The pathway to approval for most drugs for human use involves pre-clinical (animal) trials. Lab rats are considered biologically similar to humans, pa

in terms of their social behaviour. Therefore, it is very important to not compromise the pharmaceutical trials by unnecessarily stressing the rats by harnessing the monitoring equipment. One of the technical challenges of developing an implantable system that monitors the various signals, is the relatively small size requ implantable telemetry system is a miniature transceiver implanted in an animal that senses, processes and transmits data via a wireless link to monitor vital signals of c freely moving laboratory animals. This is crucial in giving researchers flexibility and reliability, especially in studies with special experimental settings using mazes wheels and treadmills. We have plans to develop a fully implantable telemetry system for subcutaneous or intraperitoneal placement in rats that monitors th parameters as well as blood pH and chemistry, nerve activity and circadian respiratory rate rhythms. The aim of this project is to eventually develop a system with e capabilities that costs less than what is currently available, to provide more universities and researchers with the opportunity to use this technology.

In this sequel, initially, when a module of our original system was placed under the skin of a rat, the wireless link failed completely. It could not send a temperature read a centimetre! The point of failure was the commercial antenna in the module that had been designed to work outside the body, in air. Such antennas do not work unde because the electrical characteristics of skin (rat or human) are significantly different from that of air. Hence the main challenge was to design an antenna that works v placed under the skin. In addition, it was necessary to cover the module and the antenna by biocompatible material, which also affects antenna performance. Possible v of skin characteristics from one rat to another or one person to another were considered. Unlike the commercial antenna, we wanted our antenna to radiate less into th the rat/person and more away from the body because that not only increases the quality of the wireless link, maximum range (distance between implanted mo monitoring station) and battery life but also reduces the exposure of the body to radio-frequency waves. Indeed we had to consider the electromagnetic effects of fat a material around the antenna. We were able to meet all these requirements with a novel compact antenna design that is approx. one third the size of the original co antenna. We successfully demonstrated wireless telemetry transmission of temperature with the new module placed under rat skin and the monitoring station pl distance of up to 80 cm! This range is sufficient for our immediate target application supporting new medical research by Professor Paul Pilowsky's team. If necessary, further extended by increasing the power level at the monitoring station.
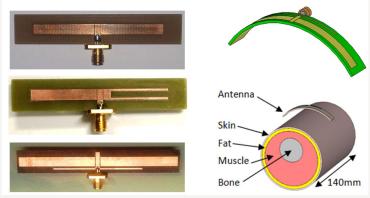
Flexible and Wearable Antennas for Bio-Medical and Healthcare Applications
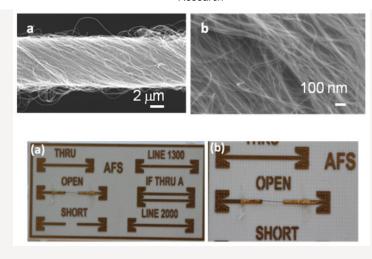
Body Centric Wireless Communication is a rapidly growing research area targeted for medical, healthcare, public safety and defense applications. The need to address transceiver specifications and real-time scenarios in close proximity to human body is continuously evolving antenna system research. Several novel miniature antennas havi dual- and wide-band operations have been designed and tested for Wireless Body Area Network (WBAN). They have significant advantages of small size, wide radiation patt the human body for maximum coverage and are less sensitive to the gap variation between human body and antenna.

A compact ultra wideband antenna is shown below with strong notch-band rejections up to VSWR = 26, that is tunable over a wide frequency range from 3.55GHz to 6.8 been designed. To estimate the stub length to notch frequency for a given interfering application, analytical expressions for the normalized stub length which is indep substrate dielectric constant is also presented. This helps to avoid hit-an-trail method and gives a good estimate of initial design parameters for notch. Proposed antenna radiation patterns and yields a measured 10dB return-loss bandwidth from 3GHz to 10.5GHz.

Characterizing Properties of Carbon Nano Tubes at Microwave Frequencies

Carbon Nanotube (CNT) yarns are novel CNT-based materials that extend the advantages of CNT from the nano-scale to macro-scale applications. We have modelled CNT potential data transmission lines. Test structures have been designed to measure electrical properties of CNT yarns, which are attached to these test structures using gold testing and microwave S-parameter measurements have been conducted for characterisation. The observed frequency independent resistive behaviour of the CNT yarn promising indicator that this material, with its added values of mechanical resilience and thermal conductivity, could be invaluable for a range of applications such as B Networks (BAN). A model is developed for CNT yarn, which fits the measured data collected and agrees in general with similar data for non-yarn CNTs.
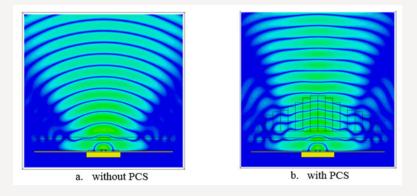
Electromagnetic Band Gap (EBG) Resonator Antennas (ERA)

We have designed, fabricated, and measured antennas based on 3D, planar and 1D EBG structures (i.e. photonic crystals). These flat microwave antennas, known as EBG antennas or Fabry-Perot cavity antennas, can give gains of about 20dB and very good efficiencies.

Enhancing Radiation Characteristics of ERAs by Improving Aperture Phase Distribution

This work focuses on achieving superior radiation characteristics of ERAs by improving their aperture phase distributions. A unique method, utilizing full-wave simula analytical analysis, has been developed to design Phase Correcting Structures (PCS) for ERAs. This method uses actual phase distribution on the physical aperture of ERAs relying on geometric optics. Several Phase Correcting Structures (PCSs) have been designed, which were later validated with the measurements of their fabricated pr Tremendous improvement in the radiation performance is witnessed in both simulated and measured results. These exciting initial results validate our proposed methodo indicate an existence of a great potential to be explored. The figure below shows the field propagation above an ERA: (a) without PCS, and (b) with PCS. It is clear that the the PCS is much more uniform and the energy is focussed towards broadside direction, thus, resulting in increased directivity.

a. without PCS   b. with PCS
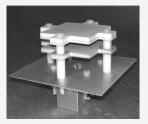
Extremely Wideband High-Gain ERAs (Gain~15-20dBi, Bandwidths >50%)

In 2014-15, we invented an innovative class of electromagnetic band-gap (EBG) resonator antennas which provide high gain and wide bandwidth with an extremely n footprint. One of the prototype developed has only 8% of the area as compared to conventional EBG resonator antennas but its performance (gain bandwidth) is a record hi for this class of antennas, while providing gain in the range of 15-20 dBi. This represents an improvement of nearly two orders of magnitude in the bandwidth compared to EBG resonator antennas. Thanks to its practical advantages of flat shape and low-cost manufacturability, it can be easily attached to a wall of a building, for example, to co building wirelessly to the National Broadband Network.
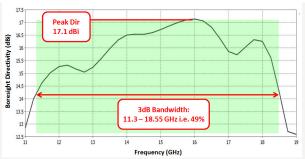
Our most recent results on such wideband high gain antennas can be found in: "Achieving high gain-bandwidth through flat GRIN superstrates in Fabry-Perot cavity antenn 2014 IEEE International Symposium on Antennas and Propagation (AP-S/USNC-URSI), pp. 1748 – 1749, Memphis, Tennessee, USA, July 6-11, 2014.

Detailed antenna design along with experimental data of another of our wideband EBG resonator antenna having composite multi-layer superstrate, is published in the pa "Wideband high-gain EBG resonator antenna with a small footprint and all-dielectric superstructures" in IEEE Transactions on Antennas and Propagation, vol. 62, no. 6, 2014
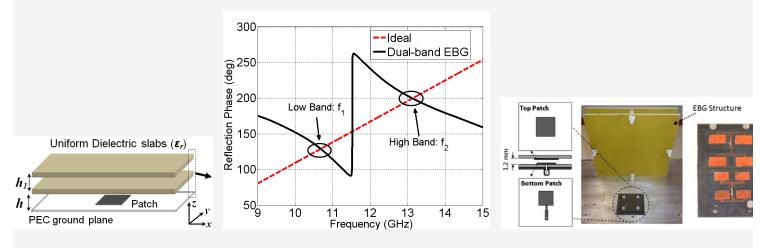
Frequency response of one of our Wideband ERA Prototypes

Simple Dual-Band ERAs

We developed a new method to obtain dual-band operation from a simple electromagnetic band gap resonator antenna. The antenna is based on a one-dimensional EBG made out of two low-cost unprinted dielectric slabs. The EBG structure is implemented as the antenna superstrate, which has been designed to provide a locally-inverted reflection phase gradient with high reflectivity, in two pre-determined frequency bands. The linearly polarised antenna design and experimental results are described in entitled: "A simple dual-band electromagnetic band gap resonator antenna based on inverted reflection phase gradient" published in IEEE Transactions on Antennas and Pro vol. 60, no. 10, pp. 4522-4529, 2012.

We have extended this concept for dual-band, dual-polarised and circularly-polarised antennas. We also designed a tri-band antenna following this concept. It needs only t cost unprinted dielectric slabs.

Low-Profile Wideband ERAs

We have designed and successfully tested wideband low-profile (thin) EBG resonator antennas. The breakthrough that contributed to this success is our design of a partially surface (PRS) with a positive reflection phase gradient. Thin single-dielectric-slab PRSs with printed patterns on both sides were investigated to minimise the PRS thickne simplify fabrication. Three such surfaces, each with printed dipoles on both sides, have been designed to obtain different positive reflection phase gradients and reflection levels in the operating frequency bands. These surfaces, and the EBG resonator antennas formed from them, were analysed theoretically and experimentally to highlight t compromises involved and to reveal the relationships between the antenna peak gain, gain bandwidth, the reflection profile (i.e. positive phase gradient and magnitude) of th and the relative dimensions of dipoles. A small feed antenna, designed to operate in the cavity field environment, provides good impedance matching (|S11|< -10 dB) a operating frequency bands of all three EBG resonator antennas. Experimental results confirmed the wideband performance of a simple, low-profile EBG resonator antenna thickness is only 1.6mm, effective bandwidth is 12.6%, measured peak gain is 16.2 dBi at 11.5 GHz and 3dB gain bandwidth is 15.7%. Please find details in a paper entitled of Simple Thin Partially Reflective Surfaces with Positive Reflection Phase Gradients to Design Wideband, Low-Profile EBG Resonator Antennas," published in IEEE Tra on Antennas and Propagation in 2012.
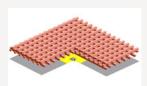
Low-Profile Dual-Band ERAs
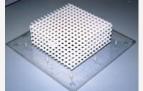
We have achieved dual-band operation in a low-profile EBGRA using a single dielectric superstate with a printed pattern only on one side. This also made use of our inverting the gradient of the PRS reflection coefficient.

Woodpile EBG Material and Antennas
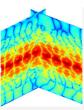
In 2003, we designed and built a woodpile 3D photonic crystal, also known as a EBG crystal, operating in microwave frequencies, and demonstrated experimentally the ex the electromagnetic band gap. The crystal is made out of cermain material.

Then we designed a planar EBG resonator antenna. Shown below, it has a resonant cavity between a ground place and a 3D woodpile photonic crystal. We employed both m and slots to feed the cavity, and investigated both linearly and circularly polarised antennas. The linearly polarised antenna design and experimental results are described in entitled "A planar resonator antenna based on a woodpile EBG material," published in IEEE Transactions on Antennas and Propagation,vol. 53, no. 1, pp. 216-223, January 2
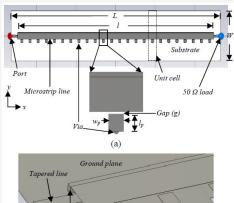
Leaky-Wave Antennas for Advanced Wireless Systems

Antenna beam steering can bring significant benefits to advanced wireless systems. Microstrip leaky-wave antennas (MLWAs) are of particular practical interest because planar low-profile configuration, ease of fabrication, and beam-scanning capabilities. In this research several planar MLWAs and arrays are developed to radiate at bores conical beam around the boresight, with simultaneous dual-side-beam scanning, dual-band forward and backward beam-scanning, and continuous beam scanning from the to the forward direction. Moreover, methods and antenna designs are proposed to steer the beam at a fixed frequency, shifting beam-steering range, and fixed-frequency bean in forward and backward directions.
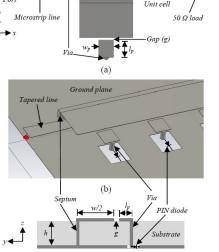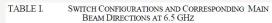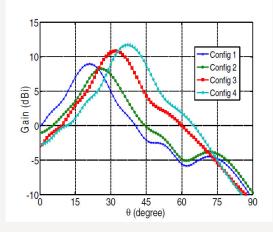
Figure 1. Proposed reconfigurable HW-MLWA: (a) top view, (b) perspective view, and (c) side view.

TABLE I.   SWITCH CONFIGURATIONS AND CORRESPONDING MAIN BEAM DIRECTIONS AT 6.5 GHZ

	Switch Configuration	Main Beam Direction (θ)
1	111111111111111111111111	21°
2	111100111100111100111100	26°
3	000011000011000011000011	32°
4	000000000000000000000000	37°
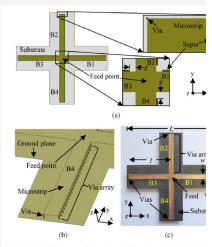
Fig. 1. HW-MLWA array design steps: (a) top view of the initial de septa, (b) perspective view of one branch of the final design with v (substrate omitted, not to scale), and (c) top view of the fabricated pro

Multi-band Dual-Polarized Shared Aperture Array

Multi-band dual-polarized shared-aperture (MBDP-SA) arrays are antenna arrays that operate in two (or more) frequency bands with dual-polarization in each band, a elements are integrated together into a common physical space by sharing the single aperture. The MBDP-SA array is of great interest in space-borne SAR system, be technique can effectively reduce the payload and size of the antenna sub-system. In this research project, main efforts are focussed on three aspects: 1) improve the specific current Dual-Band Dual-Polarized Shared-Aperture (DBDP-SA) array; 2) construct Tri-Band Dual-Polarized Shared-Aperture (TBDP-SA) array; and 3) explore some new for DBDP-SA antenna.

(a) Rear   (b) Front
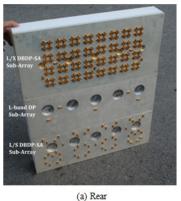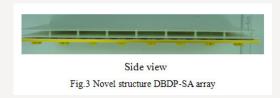
Fig.1 Improved DBDP-SA array.

(a) Rear   (b) Front
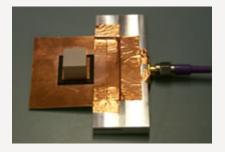
Fig.2 Photos of the TBDP shared aperture prototype array.

Side view

Fig.3 Novel structure DBDP-SA array

Compact Dielectric Resonator Antennas with Ultra-Wide 60%-110% Bandwidths

We have recently made a significant achievement in the emerging ultra wide-band (UWB) wireless communication systems, which require antennas with bandwidths gr
106%. In the past, the only way antenna engineers knew how to get such a bandwidth from a thin antenna was by removing the metal sheet underneath the antenna (known a
plane"). This is not an acceptable solution for practical systems because the lack of it allows the antenna to radiate both upwards and (unnecessarily) downwards (i.e.
electronic device on which the antenna is installed), wasting about half of its power. Saving power is crucial in UWB systems due to the severe power limits imposed by regu
2011, we made a breakthrough in dielectric-resonator (DR) antenna research, by inventing a novel DR antenna with a full ground plane and a 110% bandwidth. This disp
myth that such bandwidths cannot be achieved with full ground planes. This antenna, published in the prestigious IEEE Transactions on Antennas and Propagation in Decem
is 29% smaller but has a 30% greater bandwidth than the next-best DR antenna, which does not even have a full ground plane. Hence it is ideal for next-generation ultra-fas
and sensor applications.
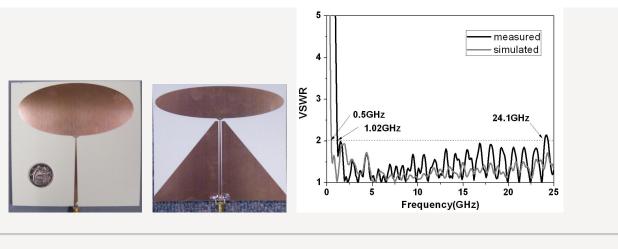
We theoretically and experimentally demonstrated that, by introducing a lower-permittivity full-length insert between the ground plane and a higher-permittivity dielectric
dielectric resonator antennas (DRA) with ultra-wide bandwidths, in the range of 60%-110%, can be designed. Furthermore, the volume of such DRAs is reduced by appro
50% using a finite planar conducting wall. Unlike vertical monopole-type hybrid UWB DRAs, these antennas radiate sufficiently in the upward direction. Unlike in printed
UWB antennas, the power radiated into the lower hemisphere is significantly less. An example prototype antenna, designed to operate in the FCC UWB band, has a dielectri
of 12 x 8 x 15.2 mm3 (or 0.124 x 0.083 x 0.157 lambda3 at 3.1 GHz), and an average measured gain of 5 dBi from 3.1 to 10.6 GHz. These antennas exploit multiple low-
with overlapping bandwidths to form an ultra-wide contiguous bandwidth. With the proposed dielectric arrangement, it is possible to efficiently couple a sufficient numbe
overlapping modes to a 50 ohm feedline using a single, simple feed.

The antenna has a remarkably small footprint of 12x8 mm2 at 3.1 GHz - the lowest frequency of the FCC UWB band. Its dielectric volume is 1459 mm3, or 1.7x10-3 lamb
lowest operating frequency of 3.1 GHz, and overall height is 15.2 mm or 0.157 lambda0. To place these results in perspective, it is worth comparing the new designs with
wideband DR designs available in the literature. To the best of our knowledge, prior to this, the widest bandwidth ever obtained from a pure DRA design is 84% . The volu
DR in that design is 0.225 x 0.172 x 0.062 lambda3 (= 2.4x10-3 lambda3) at its lowest operating frequency of 3.69 GHz. In that DRA, the DR is positioned in a non-traditi
close to the edge of an orthogonal, "vertical" ground plane, which does not block radiation towards the lower hemisphere. The widest bandwidth demonstrated by a pure DI
traditional "horizontal" ground plane, (which can be employed to shield the rest from the antenna, as discussed previously) is 78%. The dielectric volume of that design is 0.
x 0.21 lambda3 (= 5.8x10-3 lambda3) at its lowest operating frequency of 6.7 GHz.
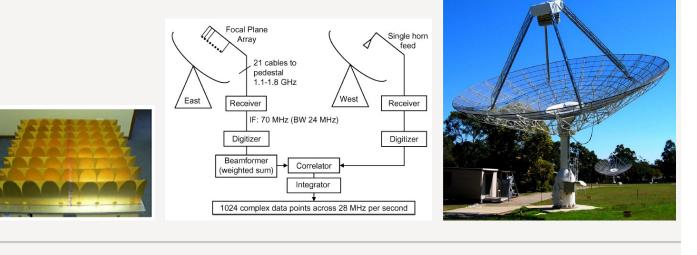
Super Wide-band Antennas

We have demonstrated that extremely wide bandwidths (ratio-bandwidths up to 1:25) can be obtained from a specially designed printed antenna with a tapered semi-ring
design is described in "A Printed Elliptical Monopole Antenna with Modified feeding Structure for Bandwidth Enhancement," in IEEE Transactions on Antennas and Pro
vol. 59, no. 2, pp. 667-670, Feb. 2011.
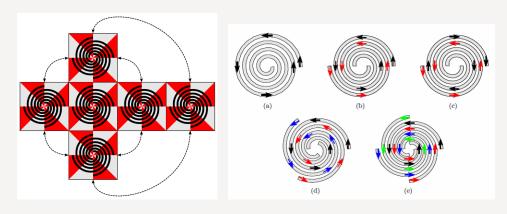
Focal Plane Arrays for Radio Astronomy

Dense focal plane arrays (FPAs) are a key technology for a new generation of Radio-telescopes. Their primary benefit is the rapid survey speed facilitated by the wide fiel provided by multiple beams. Recent advances have brought dense FPAs within reach of radio astronomy applications. A number of institutions have significant research pro this field. This technology is being considered for the Square Kilometre Array (SKA) (www.skatelescope.org). The PhD project of Douglas Hayman, conducted with CS Centre and Division of Astronomy and Space Science, investigated beamforming aspects of FPAs and evaluated their performance in Radio Astronomy. A prototype interf radiotelescope, built at CSIRO's Radiophysics Laboratory in Sydney, is used to demonstrate a suite of techniques for FPA beamforming and evaluation for this thesis. Bea solutions were experimentally demonstrated in our paper in IEEE Transactions on Antennas and Propagation, entitled "Experimental Demonstration of Beamforming Sol Focal Plane Arrays".The THEA tile, shown below, is designed by ASTRON and used for the experimental component of this research.
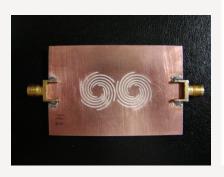
Negative Permeability of Spiral Metamaterials
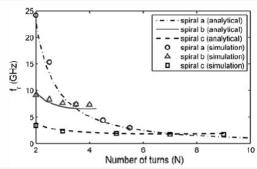
Archimedean spirals and complementary Archimedean spirals are super-compact metamaterial particles. Thanks to their convoluted geometry, unit cells can be made electri small. We theoretically analysed monofilar, bifilar, trifilar and quadrifilar Archimedean spiral metamaterial particles using point group theory and crystallography. From the s properties electromagnetic response was determined. Magnetic, electric and magnetoelectric modes of the particles were identified along with their isotropy characteristics. shown that all the particles, except monofilar spiral, are nonbianisotropic. Further, effective medium theory was applied to extract the effective permeability of the spiral med results indicated negative values for permeability in certain frequency ranges. Detailed theory and numerical simulation results are available in the paper entitled "Analysis metamaterials by use of group theory," published in the Metamaterials Journal, vol. 3, no. 1, pp. 33-43, March 2009.

Archimedean Spiral Metamaterials and Backward Waves

We have shown backward wave propagation and double negative parameters over a 19% bandwidth in a microstrip line loaded with series gap discontinuities and super complementary Archimedean spiral resonator metamaterial particles. Moreover, our equivalent-circuit model for such unit cells almost perfectly described the structu practically important frequencies (by comparison with full-wave results). We also fabricated and tested compact filter circuits with only one or two complementary spiral met particles. Our results sre summarized in the paper entitled "Backward Wave Microstrip Lines with Complementary Spiral Resonators," published in IEEE Transactions on and Propagation, Vol. 56, Issue: 10, pp. 3173-3178, Oct 2008.

We derived design equations for Archimedean spiral resonators and tested them against full-wave simulations. The details are in the paper entitled "Design of monofilar a Archimedean spiral resonators for metamaterial applications," published in IET Microwaves, Antennas & Propagation, vol. 3, no.6, p. 929-935, Sep.2009.

Frequency Selective Surfaces for Energy-Saving Glass Panel

Energy-saving glass is becoming very popular in building design due to their effective shielding of building interior against heat entering the building with infrared (IR) wave obtained by depositing a thin layer of metallic-oxide on the glass surface using special sputtering processes. This layer attenuates IR waves and hence keeps buildings summer and warmer in winter. However, this resistive coating also attenuates useful microwave/RF signals required for mobile phone, GPS and personal communication sy by as much as 30 dB. To overcome this drawback, we designed and tested a bandpass aperture type cross-dipole frequency selective surface (FSS), etched in the coatings o saving glass to improve the transmission of useful signals while preserving IR attenuation as much as possible. With this FSS, 15-18 dB peak transmission improveme achieved, for waves incident with 45 degrees from normal for both TE and TM polarizations.

Measurements and other results of this research, conducted in collaboration with the Lund University in Sweden, are available in the paper entitled "Cross-Dipole Frequency Selective Surface for Energy-Saving Glass Used in Buildings," published in IEEE Transactions on Antennas and Propagation, vol. 59, no. 2, pp. 520-525, Feb. effect of these FSSs on the transmission of infrared and visible wavelengths through energy-saving glass was investigated theoretically and experimentally in another pap Microwaves, Antennas & Propagation, Vol. 4, Iss. 7, pp. 955–961, 2010.
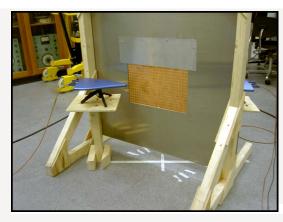
Switchable Frequency Selective Surfaces to Reconfigure Electromagnetic Architecture of Buildings

In large buildings and offices, frequency re-use methods will be required to enhance the spectral efficiency and capacity of wireless communication systems. This observatio to the concept of electromagnetic architecture of buildings. Passive bandstop FSSs can be used to enhance the electromagnetic architecture of a building, and hence to spectral efficiency and system capacity, but switchable FSSs can provide a better reconfigurable solution. If switchable FSSs are placed in strategic locations of a building, th reconfigured remotely and rapidly, which is not possible with passive FSSs.

With collaborators in UK and Sweden, we designed and successfully tested a single-layer active Frequency Selective Surface (FSS) that is electronically switchable between and transparent states. It can be used to provide a spatial filter solution to reconfigure the electromagnetic architecture of buildings. The FSS measurements show that the response of the filter does not change significantly when the wave polarization changes or the angle of incidence changes up to ±45º from normal. The FSS is based on sq aperture geometry, with each unit cell having four PIN diodes across the aperture at 90 degree intervals. Experiments demonstrated that almost 10 dB additional transmission be introduced on average at the resonance frequency, for both polarizations, by switching PIN diodes to ON from OFF state.

For details, please refer to "Switchable Frequency Selective Surface for Reconfigurable Electromagnetic Architecture of Buildings," in IEEE Transactions on Ante Propagation, Vol. 58, Issue 2, pp 581-584, February 2010.

Absorb/Transit FSS

We designed and tested a novel absorb/transmit frequency selective surface (FSS) for 5-GHz wireless local area network (WLAN) applications. The novelty of the design is capable of absorbing, as opposed to rejecting, WLAN signals while passing mobile signals. The absorption of the WLAN signal is important to reduce additional multipath spread and resultant fading caused by typical reflect/transmit FSSs. Our FSS consists of two layers, one with conventional conducting cross dipoles and the other with resis dipoles. The FSS has good transmission characteristics for 900/ 1800/1900-MHz mobile bands and performs well for both horizontal and vertical polarizations.

Later we modified the FSS to obtain even better performance, for example, for both horizontal and vertical polarizations at oblique angles of incidence. The distance betwee layers has been successfully reduced to one eighth of free-space wavelength. This small distance makes it more compact as compared to the conventional Salisbury screen achieving an acceptable absorption in the stopband.

The details of our designs and test results can be found at "Oblique Incidence Performance of a Novel Frequency Selective Surface Absorber," in IEEE Transactions on Ante Propagation, Vol. 55, no. 10, pp. 2931 – 2934, Oct. 2007, and "A Novel Absorb/Transmit FSS for Secure Indoor Wireless Networks with Reduced Multi-path Fadin Microwave and Wireless Component Letters, Vol. 16 (6), pp. 378 - 380, June 2006.
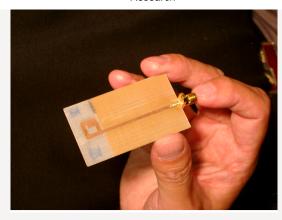
Dual-Band Artificial Magnetic Conductor (AMC) Surfaces

AMC surfaces have many advantages and interesting properties due to their unique reflection characteristics, with near zero reflection phase. We have designed and pr novel dual-band AMC surface, which has a very wide upper AMC band and a narrow lower AMC band, and therefore suitable for multi-band wireless/mobile applications.

Fully Printed Quad-band Antennas for Wi-Fi IEEE802.11 and other WLAN Applications

Our fully-printed antennas can be fabricated and integrated to WLAN systems at almost zero cost by printing them on the same circuit board (e.g. FR4) with the radio circuit same standard fabrication methods. They are extremely compact: an antenna with a radiating element of 1cmx1cm covers all four IEEE standards (802.11a, 802.11b, 80 802.11n) as well as HiperLAN2 with a VSWR less than 2. We have also developed a packaging solution where the rest of the circuit can be shielded to satisfy EMC regulati leaving the antenna (on the same board) open for radiation. The advantages of the antennas based on this technology are: Lightweight; Radiates almost every direction in shadow region); Microstrip and co-planar waveguide (CPW) designs available; Compatible with all printed microwave circuits, including stripline circuits; Excellent because no cables, connectors, soldering or any mechanical attachments are required to connect the antenna to the radio; No protruding parts that are likely to break; Covers WLAN bands with one antenna (e.g. Wi-Fi IEEE 802.11 a, b, g, n, HiperLAN2 etc.); Excellent matching, i.e. input reflection < -10 dB (VSWR < 2) in all WLAN bands (e. GHz, 4.9-5.1 GHz, 5.15-5.35 GHz, 5.725 – 5.825 GHz); High efficiency (~60-70% on FR4; can be increased to over 90% on Duroid.); Ideal for internal mounting in a c device, to achieve very wide beam coverage; diversity reception or MIMO arrangements available; can be dsigned to cover other multiple bands (e.g. multiple mobile/c bands) and other applications.

How can I optimise a UWB System to operate well over a range of directions?
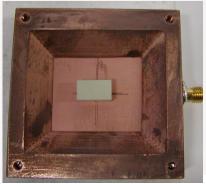
Emerging ultra-wideband (UWB) communication devices will need to operate well, not just in one direction but over a range of directions. However previous UWB system optimisation techni[...] been limited to one direction only. A system optimised in one direction may not work well in other directions due to pattern instability, or direction-dependent transfer function, of the an[...] developed the concept of frequency-domain correlation patterns and proposed a figure of merit, the pattern stability factor (PSF), which can be determined from simulation of experimen[...] With these new concepts, we have demonstrated the optimisation of UWB systems to operate well over a range of directions. The concepts and their applications are available in the PhD [...] papers by Tharaka Dissanayake.
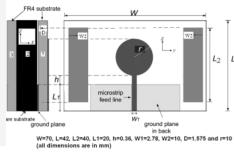
Antenna Techniques and Technologies for Ultrawideband (UWB) Systems
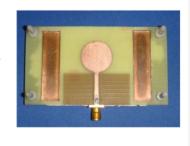
We developed antenna solutions for emerging UWB systems with very high data rates. Our research includes investigation of band-notching techniques to reduce UWB inference wi[...] wireless systems and studies of antenna dispersion and pattern stability in the time domain.

High Gain Antennas with Planar Surface-Mounted Short Horns

In collaboration with University of Delhi, we have developed compact rectangular dielectric resonator antennas with surface mounted planar horns for gain enhancement. One configuration [...] an aperture-fed rectangular dielectric resonator antenna and a planar surface mounted horn. We have achieved 10 dBi gain and a good efficiency from this configuration. The surface mou[...] increases the gain of the standard dielectric resonator antenna by 4.9 dB. Total height of the prototype antenna shown in the figure is only 0.172 lambda and the aperture size is 0.96 lamb[...] antennas can be easily adapted to low-profile, high-gain and array applications.
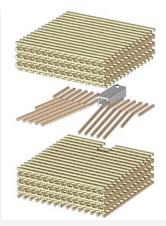
Photonic Crystal/EBG Based Horn Antennas

Waveguides and antennas made out of dielectric, as opposed to metal, are expected to perform better at Terahertz frequencies, as they do not suffer from the skin-effect loss of metal[...] investigated horn antennas, horn arrays, waveguides, bends and junctions made out of 3D woodpile photonic crystals, which can be implemented at THz frequencies in high-resistivity S[...] other materials. Our concepts have been tested by fabricating and measuring scaled-up prototypes operating in microwave frequencies. We have also designed a special broadband couple[...] a photonic crystal waveguide to a rectangular waveguide or a coaxial cable, for example for testing using a vector network analyser. We demonstrated > 6% bandwidth both theore[...] experimentally.

Broadband Microstrip Patch Antennas for Wireless Computer Networks

We have designed and tested broadband, compact E-shaped patch antennas for wireless communication networks, operating in frequencies from 4.9 GHz to 6.0 GHz. We made two
achievements. First, we designed an E-shaped antenna, which is intrinsically compatible with a printed microstrip circuit. It can be made out of a single sheet of metal and can be mou
microstrip circuit without expensive coaxial connector. Second, we designed a unique E-shaped antenna with corrugation to reduce the width of antenna. We successfully miniaturised the a
fit it inside a thin (4mm) PC (or PCMCIA) card extension. In fact, we were able to squeeze in two of these antennas into a single PC card of standard width (54mm), for diversity communica
achieve excellent isolation (>20 dB) and matching (<-10 dB) over the whole WLAN frequency band.

Theoretical Analysis of Photonic Crystal Structures

 We have developed and successfully implemented, in both personal computers and supercomputers, efficient theoretical methods to analyse and design complex electromagnetic (micro
optical) circuits based on photonic crystals (PC). Among our novel techniques is a PC-based Perfectly Matched Layer (PML) absorbing boundary for use with the finite difference time doma
method in the analysis of waveguides in 3D photonic crystals.

We have applied these techniques to model wave propagation in various guiding structures such as bends, junctions and tapers in 2D and 3D crystals. From this analysis, we can
transmission and reflection coefficients, propagation characteristics, phase and delay response, etc., of a component or a system.

New Closed-form Green's Functions for Microstrip Circuits and other Layered Structures

When combined with the spatial-domain Method of Moments (MoM), our new closed-form functions now enable efficient and accurate analysis of microstrip circuits and antennas with
approximations. The key feature in this new MoM is that the four-dimensional integrations in MoM matrix elements can be solved analytically, completely eliminating the need for expensive
integrations. Our new closed-form functions are simpler and more flexible than previous such functions, and (unlike previous ones) they do not require additional (Taylor series
approximations. The computer time required for the analysis of an example microstrip line using the new MoM was three times less than the next best method, which required some
integrations.

Integrated-design of Hybrid-resonator Antennas for Broadband Wireless and other Communication Systems

 Shown below is the first baby of our recent project on broadband hybrid-resonator antennas. This first prototype of a Dielectric Resonator on Patch (DRoP) antenna demonstrated a ba
24%. This design also proved, both theoretically and experimentally, that the electromagnetic fields in a dielectric resonator can be efficiently coupled to the fields in a patch resona
perturbing the radiation characteristics of each resonator.
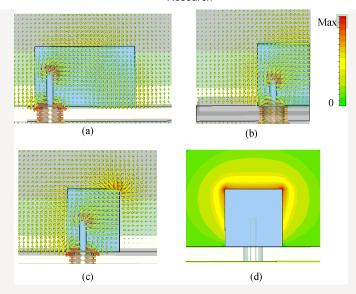
Singularity-enhanced Finite-different Time-domain (FDTD) Method for Diagonal Metal Edges, Strips and Films

We developed the world's first and still the only singularity-enhanced FDTD method for metal edges not parallel to the grid. The edges were assumed to be diagonal to cell faces. Compare

conventional spit-cell model, the computer memory required for an FDTD analysis of a structure could be reduced by up to 27 times and the computing time could be reduced by up to

without sacrificing the accuracy of results. On the other hand, when the same grid was used, the accuracy of results improved by a factor of more than 3 compared to the split-cell model, a

of more than 7 compared to the staircase model. The new equations were stable in all tests, and even in most demanding tests when the computational speed was further maximised by

the time step (Dt) to the maximum allowed by the FDTD method! (i.e. stability factor of 1!)

The key to this success was the derivation of enhanced FDTD equations for nodes near the edges by considering rigorously the singular nature of the electromagnetic fields. The table here

improvement of accuracy achieved by using the enhanced FDTD equations. The new equations are simple to implement in a standard FDTD code. They are ideal for the analysis and

microstrip components and high-speed digital circuits where thin metal films or strips with diagonal edges are encountered.

Low-profile Dielectric-resonator (DR) Antennas

We designed and tested the world's first low-profile, circularly polarised, rectangular dielectric-resonator antenna (DRA) in 1995. The radiating element of this antenna is shown in the photo

have also designed many other dielectric-resonator antennas, including a low-profile linearly polarised DRA, for various applications (see publications). We pioneered the FDTD analy

antennas and published the first radiation patterns of a DR antenna obtained using the FDTD method in 1995.

(a)

(b)

(c)

(d)

Australian Antenna Measurement Facility (AusAMF)

AusAMF is operated by a consortium of Australian Universities, Industry and CSIRO. The facility provides access to a shielded anechoic chamber offeri
spherical near-field measurement capability for small antennas operating in the frequency range of 1-20 GHz, primarily for research purposes.

The facility was established under an ARC LIEF (Australian Research Council Linkage-Infrastructure Equipment and Facilities) grant and is hosted by C
ICT Centre located in the Sydney suburb of Marsfield.

The facility currently consists of a spherical near-field turn-key NSI-700S-50 system within a 6m x 3.3m x 3.3m anechoic chamber. It has been designed to operate up to fre
of 20 GHz, and uses an Agilent PNA (E8362B) as a receiver. The measurement system is capable of supporting a ~40kg antenna under test (AUT) up to a diameter of approximately 1.2m.
the standard gain horns and the probes available at AusAMF are as follows:

Waveguide Band	Frequency (GHz)	Probe	Standard Gain Horn
WR975	0.75 - 1.12	x	x
WR650	1.12 – 1.70	x	x
WR430	1.70 – 2.60	x	x
WR340	2.20 - 3.30	x	x
WR229	3.30 – 4.90	x	x
WR159	4.90 – 7.05	x	x
WR112	7.05 – 10.00	x	x
WR90	8.20 – 12.40	x	
WR75	10.00 - 15.00	x	x
WR51	15.00 - 22.00	x	x

Smart Surfaces: Large Area Electronics Systems for Internet of Things Enabled by Energy Harvesting

9 authors, including:

Luca Roselli
Università degli Studi di Perugia
273 PUBLICATIONS **2,132** CITATIONS

SEE PROFILE

Federico Alimenti
Università degli Studi di Perugia
200 PUBLICATIONS **1,263** CITATIONS

SEE PROFILE

Paolo Mezzanotte
Università degli Studi di Perugia
144 PUBLICATIONS **1,211** CITATIONS

SEE PROFILE

Giulia Orecchini
Università degli Studi di Perugia
24 PUBLICATIONS **347** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Development of ultra wide band analog predistorsion of LASER diodes for Radio over Fiber (RoF) systems View project

Project   ADAHELI View project

Smart Surfaces: Large Area Electronics Systems for Internet of Things Enabled by Energy Harvesting

This paper focuses on large-area "smart surfaces," RFID systems, and wearable RF electronics that could substantially benefit from multisource energy harvesting.

By Luca Roselli, *Senior Member IEEE*, Nuno Borges Carvalho, *Senior Member IEEE*,
Federico Alimenti, *Senior Member IEEE*, Paolo Mezzanotte, *Member IEEE*,
Giulia Orecchini, Marco Virili, *Student Member IEEE*,
Chiara Mariotti, *Student Member IEEE*,
Ricardo Gonçalves, *Student Member IEEE*, and Pedro Pinho, *Member IEEE*

ABSTRACT | Energy harvesting is well established as one of the prominent enabling technologies [along with radio-frequency identification (RFID), wireless power transfer, and green electronics] for the pervasive development of Internet of Things (IoT). This paper focuses on a particular, yet broad, class of systems that falls in the IoT category of large area electronics (LAE). This class is represented by "smart surfaces." The paper, after an introductory overview about how smart surfaces are collocated in the IoT and LAE scenario, first deals with technologies and architectures involved, namely, materials, antennas, RFID systems, and chipless structures; then, some exemplifying solutions are illustrated to show the present development of these concurrent technologies in this area and to stimulate further solutions. Conclusions and future trends are then drawn.

I. INTRODUCTION

Looking at the telecommunication (TLC) market development worldwide in the last five years, we see an average decrease of about 5% per year, in both gross domestic product (GDP) contribution and employment [1]. Against this trend, the global information and communication technology (ICT) market has remained grossly constant. Some compartments, in fact, are experiencing an opposite trend. Beside some sectors related to new consumer products, such as tablets and smartphones, new Internet of Things (IoT) related products are growing at two digits per year and some estimations from big players report a market value in the order of trillions of dollars in the next decade [2], [3].

The vision behind IoT is in fact to connect objects directly to the Internet so as to allow them to provide information directly to the web without any human intermediation [4]. This vision will have a great impact on several electrical and electronic technologies, ranging from the basic physical layers (technology platforms) to the highest ones: communication protocols, software, human interface, and information management.

In terms of physical layer, IoT actually inherits the technologies developed for wireless sensor networks (WSNs), bringing the distributed nature of that to the extreme. WSN, in fact, can be synthetically seen as a mesh of sensor nodes purposely conceived to build and communicate a sensed image of a monitored area. Within IoT, the sensing nodes are not purposely deployed to monitor something specific; instead, they are simply hosted by "objects" in order to provide the information they are able to collect and make it available on the web. Conceptually speaking, there is not a big difference, but in terms of related technological challenges, there is. Nodes must be hostable by objects; objects are of many types, and the better the nodes fit themselves transparently to the hosting objects, the denser and more reliable the information they are able to provide.

Several challenges can be envisioned in this evolution: first, electronic systems must be mechanically flexible, thin, and miniaturized in order to conform to the shape of as many objects as possible; second, the adopted materials have to be as recyclable as the hosting objects in order to avoid pollution from guest apparatuses; last but not least, hosted nodes must be autonomous, because they cannot be either connected to the grid or powered by life batteries.

A great technological paradigm shift is going to be pulled by the development of IoT; green materials, autonomous systems, ultralow-power circuits, energy saving protocols, and energy harvesting (EH) are concurrently mandatory.

Conversely, this technological evolution is pushing new solutions and architectures, so far constrained by the limits of conventional technologies. The development of inkjet printing techniques, the introduction of very cheap and eco-friendly materials compatible with roll-to-roll (R2R) circuit realization techniques [5] and so forth, open new horizons also to large area electronics (LAE).

LAE, first applied to printed photovoltaic and organic screens [6], is actually at the onset of its development. The development of R2R techniques and related materials, in fact, is allowing for tremendously increasing the dimensions of LAE circuits and systems, from the present tens of centimeters to meters and beyond. Along this evolutionary scenario, new configurations and architectures, based on massive integration of large circuits over conformable surfaces, can be envisioned, opening the way to what can be called the smart surfaces (SSs) approach.

SS, in turn, can be seen as a branch of IoT evolution. Large 2-D arrays of autonomous sensor nodes, for instance, make possible granular tracking of whatever parameter, ultimately providing augmented imaging of the environment; large 2-D arrays of tags can provide a very low-cost platform for precise localization and location-based services (LBS), thus enabling the realization, for instance, of smart floors [7] and smart wallpapers [8]; inheriting quasi-optical approach [9] contact-less electromagnetic (EM) wave processing (filtering, frequency conversion, selective shielding, etc.) can be conceived even at low frequencies.

On the one hand, SS is thus an approach stimulated by technologies pulled by IoT; on the other hand, it contributes to a class of architectures, within LAE, that can even widen the huge horizons of IoT applications in a sort of virtuous circle.

According to this wide vision, the paper is organized as follows. A review of the technologies suitable for distributed systems implementation to a large extent is given in Section II; then, the explanation about how radio-frequency identification (RFID) can be considered one of the most suitable technologies for the implementation of IoT architecture, and how it can be naturally integrated with SSs, is provided (Section III). In order to deal with the implications of the development of RFID systems compatible with SS and LAE evolution of IoT, two specific sessions have been provided: the first relates to antenna implications (Section IV), and the second relates to electronic architectures for RFID tags (mainly chipless ones; Section V). In order to smoothly bring the reader from relevant technologies to applications, quasi-optics, as an example of general approach concurrently exploiting the mentioned technologies to provide a platform suitable for LAE, is described in Section VI. After this, some application examples of how SS concept can be articulated, according to the IoT paradigm, are described; specifically, smart floor (Section VII), smart shoes as useful subsystems for the implementation of smart floors (Section VIII), and energy skin (Section IX). In order to have a vision of the logical links behind the many topics dealt with in this paper, a synoptic picture can be found in Fig. 1.

II. TECHNOLOGY FOR LAE DISTRIBUTED SYSTEMS

Given the described scenarios of SS, it is clear that the technologies involved have to be compatible with LAE
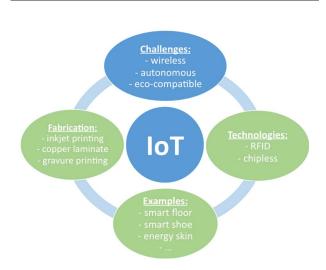
Fig. 1. *Synoptic view of the paper structure reflecting the IoT vision described in the Introduction.*
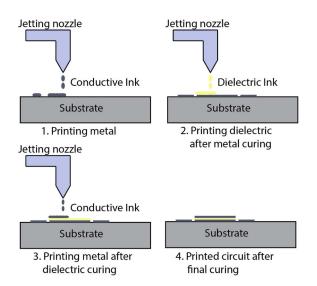
Fig. 2. *Inkjet printing technique, synoptic description. The ink is deposited on the substrate with the desired pattern and after a curing procedure the next layer can be printed. The thickness of each film depends on the number of passes.*

approach, as described in the Introduction; according to [6], for example, the eco-compatibility of the materials, the mechanical flexibility of the substrates to make the device conformable to the hosting surface, and the life cycle (operability and biodegradability) have to be accounted for.

The new techniques proposed from different research groups [10]–[12] focus mostly on the printing techniques that can be industrialized easily with R2R methods, already investigated for the traditional printing on paper. Within the printing methods, one of the most attractive and investigated in the last decade is the inkjet printing of conductive and dielectric layers (Fig. 2). This technique allows circuits to be printed on almost any kind of substrate: from the photographic paper to liquid crystalline polymers (LCP) or Kapton, that are flexible, thin, and eco-compatible; to glass or poly methyl meta acrylate (PMMA, commonly referred to as Plexiglas) that are usually thicker but not flexible.

A first proof of concept for the inkjet printing technique at radio frequency (RF) and microwaves has been obtained on photographic paper by printing simple structures, such as antennas and other planar circuits and devices [10], [13]–[22], with a technology stack simply composed by the substrate and the conductive layer (usually realized by means of nanoparticle silver ink).

Recently, this method has been improved thanks to the development of new inks, usually composed by a polymer and a solvent (i.e., SU-8, PVP, PMMA, PEDOT, etc.); more complex fabrication procedures of multilayer structures are now possible, as shown in [23] and [24]. Another noticeable feature of this technology is represented by the resolution: conductive tracks of 50 μm of width and space gaps of 50 μm can be printed, allowing the design of

millimeter-wave (mm-wave) frequency devices [25]. The combination of the quite high resolution (considering the simplicity of the technology) with the multilayer featuring also gives the opportunity to manufacture very easily matrix of passive and active devices that can be used in the SS development. It is worth noting that the process can be developed in a few steps based on the design and mostly on the number of layers needed. For example, a metal–insulator–metal (MIM) device can be realized by printing silver on the substrate [26], curing it in the oven in order to create a surface that, with the proper treatment (for instance, UV–Ozone exposure or preheating), is ready for the insulator printing. Then, the dielectric is cured as well and the last metal layer can be printed on top of it. The entire structure can be then cured at temperatures ranging between 130 °C and 200 °C.

In a perspective of industrialization of the inkjet printing technology, the hypothesis of R2R manufacturing of circuits is being investigated and some examples are already reported in literature for solar cells and other devices [27], [28].

Today, many pros have been mentioned for the inkjet printing method, and others can be found in the non-use of wasted chemicals (as is for the traditional lithographic technologies), in the no-need of clean-room environment for the fabrications, in the low-cost and rapidity of manufacture, and in a R2R perspective compatibility. However, the necessity of a curing procedure after the printing of a layer still represents the biggest inkjet limit for two main reasons: first, the time and type of curing are dependent on the inkjeted materials and on the material stack-up; second, especially on a laboratory level, the curing can cause misalignment issues, given the fact that the sample is removed and then replaced in the printer after the last layer curing.

Currently, the platform mostly used worldwide by researchers is the Dimatix 2800 DMP. In terms of inks, it is possible either to buy printable solutions or mix solutions in labs; for example, the combination of polymers and solvents allows layers to be printed with different electrical properties and thicknesses.

Besides the inkjet printing method, in [11], a new technique, suitable for LAE, has been proposed. It uses a metal (copper in that case) adhesive laminate technology based on the application of the standard etching process to an adhesive copper laminate material. This technique was already adopted to produce mm-wave circuits and to characterize the resulting compound substrates, as reported in [29]–[31]. A brief illustration of the metal laminate technique is here reported referring to Fig. 3.

The production process can be divided in five steps. The first step consists of the deposition of a (positive) photoresist film on the copper surface; then, the circuit layout is transferred to the photoresist by means of a photomask and ultraviolet exposure. After that, the unimpressed film of photoresist is removed using a NaOH
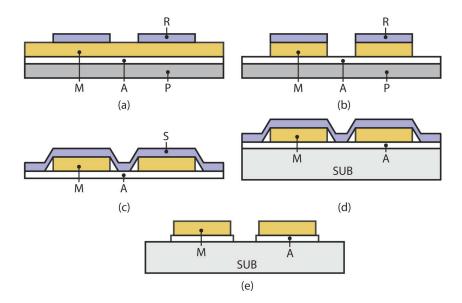
Fig. 3. *Process steps for the circuit fabrication using the adhesive metal laminate. M: metal, A: adhesive, P: protection, R: photoresist, S: sacrificial layer, SUB: hosting substrate. After the standard lithographic procedure, the copper pattern can be transferred on top of any substrate by means of a sacrificial layer and exploiting the adhesive face of the copper tape.*

developer solution [Fig. 3(a)]. In the second step, the copper tape is wet etched. As can be seen in Fig. 3(b), in this way, the adhesive layer is exposed where the copper was removed, while it remains everywhere else covered by the original copper tape that serves as protection for the adhesive underneath.

The first two steps are, in this example, similar to those adopted in standard photolithographic technology. Moreover, different ways to remove the not needed metal can be used, as, for example, by means of numerical control pattern cutting plotters.

In the third step, depicted in Fig. 3(c), a sacrificial layer is stuck on the top copper side and, finally, the protection layer on the bottom is removed. The sacrificial layer is very important because it keeps the relative distances among the layout features constant even when these are not physically connected.

The fourth step is characterized by the transfer of the etched metal to the hosting paper substrate and, finally, the sacrificial layer can be removed [see Fig. 3(d) and (e)]. The last step is also useful to remove most of the exposed adhesive material.

With this method, quasi-fully-organic circuits and devices can be realized. The performance in terms of tracks width and pitch are, at present, a bit lower than what can be obtained with the inkjet printing, with the advantages of no curing processes, the possibility to fix devices on the circuit using standard soldering techniques, and a better conductivity (the conductivity of copper laminate, in fact, is one order of magnitude higher than that of the cured nanoparticle silver ink: 5.8×10^7 S/m versus about $6 \times 10^6 - 1 \times 10^7$ S/m obtained with at least five layers of silver).

To verify the validity of the metal laminate technology at microwave frequencies, a 50-Ω microstrip line was manufactured exploiting the Mitsubishi photografic paper as the substrate (thickness 250 μm, relative permittivity $\varepsilon_r = 3.2$, and loss tangent $\tan \delta = 0.08$). The line is 30 mm long, and the measured scattering parameters are shown in Fig. 4. The same graph also compares the performance of a similar transmission line manufactured with an inkjet printing process. The performance of the
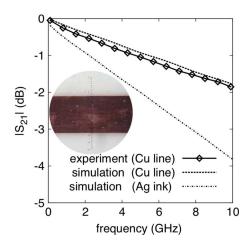
Fig. 4. *Measured scattering parameters versus frequency for a 50-Ω microstrip line on paper substrate. The line, shown in the inset, is 30 mm long; the ruler division corresponds to 17.2 μm. The graph also reports a comparison with a microstrip line of equal dimensions made with an inkjet printing process (Ag ink, 3-μm thickness, $\sigma_{ink} = 1.1 \times 10^7$ S/m after curing). After [11].*
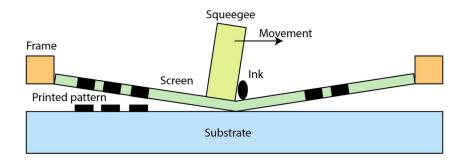
Fig. 5. *Screen printing technique. The squeegee is used to press the screen with impressed pattern and transfer it on top of the substrate.*

metal laminate structure is superior compared to that of the inkjet printed line. At 10 GHz, the measured specific losses for the metal laminate microstrip are about 0.6 dB/cm [11]. Results at higher frequencies are also reported in [31]; in particular, an insertion loss of about 1.8 dBm/cm at 30 GHz is demonstrated.

The tradeoff is the adoption of a photolithographic step; it is worth mentioning, however, that the process is still compatible with R2R implementation.

Other possibilities for LAE are the screen printing [12] and the gravure printing [12], both of them adaptable to R2R industrialization techniques.

Screen printing consists of dragging a layer of ink across the surface of a screen and squeezing it through the open pores of the patterned mesh onto the substrate (Fig. 5). In general, the thickness of the printed layer and the achievable resolution depend on the density of the mesh and on the ink properties. Usually, the ink viscosity is in a range of 1–50 Pa $\times$ s, and this allows for printing with a resolution of about 100 μm and a thickness up to 100 μm. Until now screen printing has been adopted to realize low-resistance structures, solar cells, and field effect transistors (FETs), exploiting the possibility of printing very thick layers.

Gravure printing, also known as rotogravure, is considered a very high volume printing (Fig. 6) process, and it is being adopted to produce very conductive structures as, for example, capacitors, antennas, and organic devices such as pentacene-based diodes, organic light-emitting diodes (OLEDs), organic field effect transistors (OFETs), and organic thin film transistors (OTFTs).

The resolution can be of about 20 μm. The method is implemented by engraving the patterns into a metallic cylinder by a laser, by chemical etching, or electromechanically (as separate cells or intaglio trenches). Typically, the print pressure is high (1–5 MPa) in order to achieve a good ink transfer and to reduce the percentage of unprinted dots caused by the surface roughness of the flexible substrate. However, this high pressure of printing makes this technique suitable only for robust substrates with no soft, previously printed, layers that could be damaged.

To summarize, the technologies described in this section are characterized by some common features, such as: flexibility and conformability, compatibility with large area realization of circuits and interconnections, use of additive or mixed (subtractive/additive) deposition techniques, easy use of eco-friendly materials and, in some cases, compatibility with classical bond wiring as well as soldering techniques for electronic device placement.

Table 1 summarizes the main features of the described technologies.

III. RFID: A MIX OF CONCURRENT TECHNOLOGIES FOR SS IMPLEMENTATION

As mentioned in the outline paragraph at the end of the Introduction, after describing the technologies enabling SS development (Section II), here we illustrate RFID as a suitable means to transfer information through EM transmission between tags and readers to the Internet.

An RFID tag has a unique identification code and a memory used to store information, while the reader can
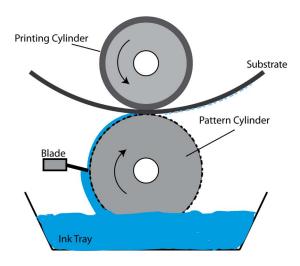
Fig. 6. *Gravure printing process. Note that these machines can be from 20 cm to 3 m wide with a diameter ranging from 5 cm to 5 m.*

Table 1 Technologies Comparison

Technology	Resolution	Cost	Waste	Speed
Inkjet	$50\,\mu m$	Low	Low	Fast
Copper	$100\,\mu m$	Low	High	Medium
Gravure	$20\,\mu m$	High	Medium	Very fast
Screen	$100\,\mu m$	Medium	Low	Very fast

write or read data on and from the tag through RF transmissions. The tag is usually attached to an object that needs to be identified and monitored, or contains information to be read. In typical RFID applications [32], the main goal is to provide mapping of physical objects that are equipped with an RFID tag attached or embedded. In this way, RFID technology inherently leads to identifying, tracking, and localizing, but it can also enable the storage and transmission of information regarding object status and surrounding environments by embedding sensing capabilities.

As mentioned previously, IoT is most commonly described as a structured system of technologies concurrently working to capture meaningful data from objects and communicate them through networks to a decision-taking level. Useful keywords, whenever a definition of IoT is given, include: smart objects, autonomous logistics, machine-to-machine communication, RF technologies, centralized information, and so on.

In order to implement these functionalities, an IoT platform, an SS for instance, has to provide at least unique object identification together with wireless communication for data management.

The IoT physical layer can thus be naturally thought of as meshes of RFID and sensor nodes. The RFID tag that embeds a sensor can use the same working principles and protocols of the conventional RFID tags; a differentiation can be made, considering the way sensor tags are powered up, in active, semipassive, and passive tags; actually, the difference consists of the way they are powered.

Active tag sensors may use customarily conceived communication protocols or rely on RFID standards enabling the tag to be easily integrated into the existing wireless infrastructures so that they will not require expensive readers. In [33], a system architecture was developed to integrate WSNs and the RFID systems. Bluetooth and ZigBee technologies are adopted as the communication protocol of WSNs to meet the requirements of a large number of sensor nodes, large areas, and low cost.

Active tags with integrated sensors are used in several applications, including temperature and position monitoring, vibration detection, blood pressure, heart beat monitoring, and more. This type of sensor tags, besides having larger amount of energy provided by batteries, affords both a large range and multiple functionalities; some of them have also external buses that enable the use of connected external sensors [34].

Passive tags with an integrated sensor operate without battery, collecting the needed operational energy from external environment sources. The main requirements of this class are high energy conversion efficiency, large storage capability, and overall low-power consumption. Passive systems have usually low operating range and limited processing capability.

In this field, research activities are directed toward ultralow-power design of integrated sensor tags [35], antenna design for improving reading range [36], sensor overall performance optimization [37], and development of optimum protocols allowing for additional power saving [38], [39].

Typical applications are, for instance, temperature monitoring [40], photodetection [41], motion detections [42], etc.

Semipassive tags use both batteries and energy sources coming from the environment; they can integrate more operational capabilities than the passive ones, since they can exploit a higher amount of energy. For the same reason, semipassive RFID tags are more suitable to integrate a sensor; the operational methodology is similar to the passive one, using the reader signal to interrogate and cause a response from the tag. The primary difference is that the semipassive tag does have a battery used not to generate a response, but only to power electronics, like sensors incorporated in the tag, while exploiting RF circulation to reply to the reader interrogation. Just like passive tags, semipassive ones are limited in terms of slow reading speeds and short reading distances. In [43], the design, realization, and experimental validation of a battery-assisted RFID tag with sensing and computing capabilities, conceived to explore heterogeneous RFID-based sensor network applications, was presented.

Dealing with battery-less devices that are, by far, the most interesting ones for IoT, LAE, and SS applications, we can now focus on the typical energy sources adopted. No battery devices can exploit solar scavenging [44], [45], vibrational [46], [47], or RF radiations [48]–[52].

In many cases, wirelessly powered systems can harvest energy from the incident RF waves, thus generating the required direct current (dc) voltage to power the system. The EM energy present in the environment may be significant in some regions (some urban areas or indoor sites); here the use of proper conversion devices may allow enough energy to be scavenged. RF power harvesters usually consist of an antenna, an impedance transformation network, a rectifier, and a storage element. In [53], a design example of integrated rectifier antennas (rectennas) for wireless powering at low incident power densities, from 25 to 200 $\mu W/cm^2$, is given.

Regarding mechanical energy scavengers, they may be categorized upon the kind of energy source exploited; there are scavengers providing energy from a constant motion over extended periods of time (e.g., turbine air flow) or others based on intermittent motion (e.g., human

step). Particular attention is given to vibrational energy sources; in this case, the amount of energy that can be harvested depends mainly on the vibration amplitude and frequency and on the vibrating mass (mass of the harvesting device) [54]–[56].

In order to power the systems by using energy sources available in the working environment, proper energy transducers thus have to be chosen. Focusing on distributed systems (IoT, LAE, and SS), it is worth referring to typical parameters for optimum transducer design since, in these applications, the different kinds of energy available in the various scenarios may lead to the need of integrating more than one energy conversion system into the same device, in order to take advantage of all the accessible energies. The main design parameters are the available power output, the electrical impedance, and the operating voltage. Table 2 summarizes them for the most common energy transductions [57], [58].

To conclude this section, the mentioned reasons why RFID can be considered a useful technology for the implementation and development of LAE and, more specifically for SSs, are summarized. RFID systems, especially in the passive version of tags, are communication systems characterized by the following characteristics: compatibility with very low-power communication protocols, compatibility with the constrains posed by the hosting objects, especially when they are implemented by adopting the technologies described in Section II, inherent capability to gather information from the environment (sensor tags) and from the hosting objects, a low number of electronic devices (inherent circuital simplicity) and, finally, proven compatibility with printing technologies.

For all these reasons, RFID technology has to be considered the leading one to pursue the development of LAE and SSs as a communication platform for the evolution of IoT.

IV. ANTENNA DESIGN FOR NODES

Distributed systems for ubiquitous electronics (UE) and WSNs [59], [60] are, today, increasingly employed for monitoring and sensing applications. These systems can be seen as networks of nodes, massively distributed in the environment and communicating wirelessly with a base station. This massive deployment can be designed and structured to monitor a set of specified parameters by means of purposely conceived sensor nodes, as in the classical WSNs, or it can be obtained by integrating nodes into existing objects, which is the vision of IoT evolution. In general, the nodes can have different functionalities, however they have to provide information collected by embedded sensors (such as, for instance, identification codes, position and monitored environment parameters as well as object status) whenever it is required.

In this scenario, the RFID technology, summarized in Section III, is attractive. When RFID technology meets sensor nodes, information goes from the node to the Internet via a question-and-reply protocol between nodes, hosted by tagged objects, and readers.

Thinking about distributed systems and specifically SS, it is natural to think about radiative coupling in the available industrial–scientific–medical (ISM) frequency bands, including mainly the ultrahigh frequency (UHF) and microwave frequencies (around 0.9, 2.4, 5.8 GHz, and higher), exploiting the communication standard such as the WiFi [61]. Recently, also the ultrawideband (UWB) has been considered in order to minimize the limitations of narrowband systems (mainly localization accuracy and sensitivity to interference) [62].

Multistandard systems are often adopted for large compatibility and higher use flexibility. An example of this recent approach is given in [61] where near-field UHF nodes are combined with far-field transceivers. This allows application fields with compromised radiation capability,

Table 2 Transducer Characteristics for Several Forms of Ambient EH [57], [58]

Energy Source	Typical Electrical Impedance	Typical Voltage	Typical output power
Light	From low to 10s of KΩ	DC: 0.5 V to 5 V	Outdoors: $100s$ mW/cm^2 Indoors: $<$ $500\,\mu$mW/cm^2
Vibrational	Constant Impedance 10s of kΩ to 100 kΩ	AC: 10s of Volts	$100s\,\mu$W/cm^2
Thermal	Constant Impedance 1 Ω to 100s of Ω	DC: 10s of mV to 10 V	$10s\,\mu$W/cm^2 (20° gradient)
Radio Frequency	Constant Impedance Low KΩs	AC: Varies with distance and power 0.5 V to 5 V	Wide-range

such as in proximity of metal and liquids at short ranges, to be improved.

It is clear that antenna systems are key elements in these apparatuses. The foundation of antenna design rules can be found in [63]. The antenna design specific for distributed systems, in turn, has to take into account several instances. Beyond the typical design criteria, other aspects are: the communication standard, the form and the size in agreement with the application, the communication range as a function of the equivalent isotropic radiated power (EIRP) limits, of the environment and of the orientation, the reliability under several conditions and processes, and the material cost considering massive production [64]–[67]. Among them, the operating environment is the most critical aspect that influences the node behavior [68].

As a consequence, antenna topology and materials for SS and IoT systems, in general, have to be chosen in order to make the antennas capable of operating in variable environmental conditions and to be integrated into objects. As an example, let us think about wearable antennas for body area networks (BANs): they must be washable, flexible, and able to shield the body from radiation [69].

In general, antennas for SS should be flexible, planar, and low profile; moreover, by adopting low-cost materials and technologies (see Section II), cost per unit surface is low, thus large arrays and big elements can be adopted more easily, ultimately allowing for easy developments of LAE distributed systems and SS, even at low frequency.

Materials compliant with SS implementations must be, in turn, mechanically robust, flexible, and very low cost, while exhibiting acceptable EM performance. These materials must also exhibit high tolerance levels in terms of bending repeatability. Flexible materials, often used, are polymers and polymer-ceramics composites such as the poly-di-methyl-siloxane (PDMS) and the titanate-based ceramic [68], which provide flexibility, robustness, and resistance to harsh environments in order to protect antenna and circuit. Moreover, when eco-compatibility has to be guaranteed, recyclable materials and related technologies have to be adopted; among them, cellulose substrates in combination with inkjet printing of conductive inks [18], [70] or with conductive adhesive laminate technology [11] are increasingly frequently adopted. The aforementioned techniques, compatible with the well-known industrial printing and the R2R processes [5], are in agreement with the low-cost fabrication and are expected to facilitate widespread and very low-cost electronics for distributed systems. These fabrication technologies can also be adopted for electronic devices on flexible polymeric substrates. An example of cellulose-based antenna for smart floor applications, such as indoor localization, is documented in [71]: a low-profile planar loop antenna is realized on cellulose substrate exploiting adhesive copper tape.

Although beyond the scope of this paper, which focuses on SS, it is worth mentioning here that this evolution toward the development of more and more sophisticated "unconventional" materials for electronics has a great impact on other IoT related developments such as, for instance, implantable systems.

When applications require devices with soft visual impact (think about SS, such as smart windows, glass integrated inside and outside buildings and glass for automotive applications, or smart skins [72]), transparent and conductive materials, such as transparent conducting oxides (TCOs), and in particular indium-tin-oxide (ITO) and ITO-based multilayers, have been proposed [73]. These materials afford a good compromise between minimum electrical resistivity and high optical transmittivity in the visible light spectrum when patterned on see-through materials such as glass or transparent polymers. A technique to improve the efficiency of the transparent antennas, without affecting their transparency, is the application of a layer of highly conductive coating or metallization strips on the antenna edges where the current density is high [72].

Among SSs, or in general distributed systems within the IoT vision, a fast growing field of research is represented by wearable electronics for BANs [74]. In this case, the integration of electronic devices in the garments requires the development of wearable and washable antennas and antenna arrays on fabrics. Washable antennas can be realized by covering textile antennas by a breathable thermoplastic polyurethane coating in order to protect the device against water absorption and corrosion [69].

Although dimension constraints are not very critical for distributed systems on cheap materials, they cannot be neglected thoroughly. At RF, for example, high permittivity materials [68] and magnetic composite substrates [75] can be adopted. Artificial magnetic conductors (AMC) can be employed alternatively to the magnetic composites with similar results [76]. These approaches are simply demonstrated from the wavelength equation: wavelength is a function of the relative permittivity and permeability and adopting suitable materials to increase these values, the effective wavelength, and therefore the antenna size, decreases. Anyway, the use of these materials introduces a tradeoff between dimensions and performance: generally the substrate losses decrease the antenna radiation efficiency, thus affecting the communication range of the overall wireless system, as documented in [68], [75], and [76].

In terms of design methodology, the introduction of unconventional materials as well as the environmental constraints posed by distributed systems and systems for IoT influence and drive the design strategies. Some solutions are proposed in the literature to cope simultaneously with these constraints; in [68], shielded configurations have been analyzed to mitigate environmental sensitivity. In [69], an example of robust patch antenna for BAN application that reduces the effect of the body on the performance is illustrated; in [77], the use of broadband antennas, such as the Chebyshev monopoles with a fractional bandwidth of about 40%, is proposed to reduce the frequency

shift introduced by the background materials or by the bending effects.

In SSs, and in general in IoT applications, antennas are used both to transmit and receive information and to harvest the necessary energy from the surrounding environment. Regardless of the fact that the available energy is randomly released in the environment by existing services (GSM, WiFi, and so on, in urban and semi-urban environments [78], [79]) or purposely delivered to empower a distributed population of nodes [80], antennas have to be able to harvest EM energy. This kind of "double effect" antenna is widely referred to in the literature as "rectenna": it combines rectifying circuitry with EM to electrical energy transduction.

The fundamental codesign techniques of antennas and rectifiers, focusing on the improvement of the conversion efficiency exploiting matching circuits, are described in [53]. Harmonic-balance and load-pull simulations are often used in order to find the optimal impedances that allow best efficiency to be obtained as a function of the EM incident power.

Multiband, broadband, and circularly polarized antennas are often adopted in EH applications to reduce the overall sensitivity to the environment. An example of planar antenna array for EH applications, in agreement with the SS principle, is documented in [81]: spiral antennas are proposed in order to provide a wide operating frequency band, and the rectification circuitry is optimized and characterized for low incident power levels. Similar antenna array for UWB EH is shown in Fig. 7.

A rectenna can be used in addition to the communication antenna or, alternatively, the same antenna can be used for both EH and data transmission. The use of different antennas is generally adopted if harvesting and communication are carried out at different frequencies
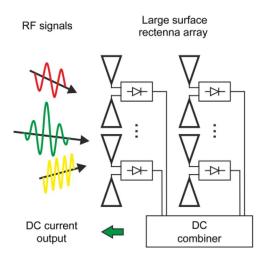
Fig. 8. *Hybrid rectenna for solar and EM–EH. The hybrid rectenna consists of a broadband antenna and a solar cell combined in the same structure. After [82].*
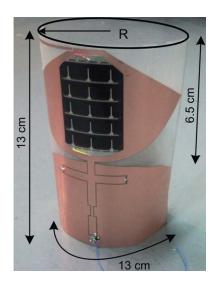
and, mainly, in order to enhance harvesting efficiency without decreasing communication performance. In any case, this requires a codesign and can affect the node dimension and complexity. Alternatively, a single antenna can be a feasible compromise if communication and harvesting share the same frequency band and if the antenna harvesting efficiency is not the main issue; examples are the passive RFID tags empowered by the reader with a standard RFID antenna able to provide both EH and communication capability, as documented in [64] and [65].

However, the EM–EH is not the only solution to energize nodes and SS elements: hybrid EH, in fact, can be performed by exploiting different energy sources; that is, combining the rectennas with other sources of energy.

Recently, a low-cost and conformal structure for hybrid EH, based on low-cost and flexible poly-ethylene terephthalate (PET) substrate and an amorphous silicon solar cell, has been proposed in [82]; other examples are discussed in [20] and [83]. The use of low-cost, printable photovoltaics deposited on flexible substrates to form part of the antenna radiating structure is documented in [84]. In all these contributions, the main challenge appears to be the codesign of the antenna and solar cell in order to avoid reciprocal detrimental effects. The antenna is designed with a solar cell integrated on top of the radiating surface to obtain good radiation characteristics and to incorporate connections for the extraction of the dc provided by the solar cell. In [82], it was demonstrated that, by adopting thin film solar cells and paying attention to not exceeding the perimeter of the RF antenna, the solar cell can be integrated so as not to significantly affect the performance of the antenna. The hybrid rectenna for solar and EM–EH documented in [82] is shown in Fig. 8.

Fig. 7. *Scheme of the rectenna array able to collect RF energy even in a broadband (depending of the kind of receiving antenna element), according to the solution adopted in [81].*

In terms of industrial assembly, the use of new materials and the need for low-cost and large-scale fabrication techniques would be fostered by architectures more insensitive against geometrical errors and placement misalignments. Following this direction, EM coupling can be exploited instead of ohmic contact to connect antennas on a flexible substrate to Si chips or, more generally, to the active circuitry [85]. This technique exploits a planar heterogeneous transformer, with the primary and secondary windings implemented on the antenna substrate and the Si chip, respectively. In this way, the galvanic contacts and the soldering process are avoided, and the final structures have been proven much more robust to alignment errors than traditional ohmic techniques without any significant performance degradation. By artificially applying a misalignment of 150 μm, which is the common dimension of a soldering pad, the technology proposed in [85] causes a loss increment of only 0.1 dB, due to the applied misalignment, while the same misalignment, when adopting ball grid arrays, would have implied missing the contact. This approach is quite attractive especially in combination with nonrigid materials where mechanical stress can result in significant substrate deformations. Recent contributions show the application of the aforementioned approach to inkjet printed antenna on PET substrate coupled to the chip for wireless nodes [86] and a textile patch antenna, magnetically coupled to the active circuitry, suitable for garment integration (see Fig. 9) [87], the reflection coefficient of which is shown in Fig. 10.

In order to testify the robustness of the proposed solution against uncertainties inherent to textile and garment electronics assembly, the radiation efficiency of the antenna, as a function of horizontal and vertical misalignment of the two transformer windings, used to perform the magnetic coupling, is plotted in Fig. 11(a) and (b), respectively. This technique can be extended also to interlayer communications of multilayer devices, and an example is documented in [88], where the circuitry fabricated on different garments communicates by means of overlapped coils.

V. CHIPLESS APPROACH

It has already been discussed that SS represents a possible technology for the development of killer applications within the IoT paradigm. Thanks to the convergence of several technologies, ranging from EH and sensing to RFID and LAE, SS can afford complex functions. For example, SSs can be used in intelligent buildings for structural monitoring and alarms (fire, for instance). To achieve these goals, SS must be equipped with a huge number of tags, each one monitoring a small portion of the surface itself.

The passive RFID systems can be divided into two families, namely, chip-based and chipless tags. The first family usually exploits a low-power complementary metal–oxide–semiconductor (CMOS) chip to implement the

Fig. 9. *Example of textile patch antennas operating at 2.4 GHz. The antenna and the ground plane are made of Flectron, a conductive textile, and the substrate is black foam. These materials are flexible and integrable in the garments. Magnetic coupling is exploited to connect the antenna and the active circuitry on flexible substrate. After [87].*

main tag functions [RF carrier rectification, dc voltage regulation, amplitude shift-keying (ASK) demodulation, IDentification code (ID) decoding, load modulation, etc.] and thus can be adapted to sensor applications by a few additional circuit blocks [signal conditioning,
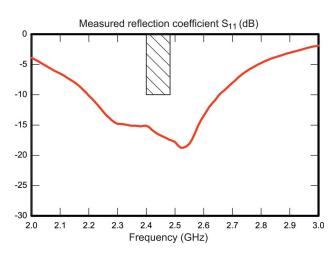
Fig. 10. *Measured reflection coefficient of the previous textile antenna in the frequency band 2.4–2.48 GHz. After [87].*
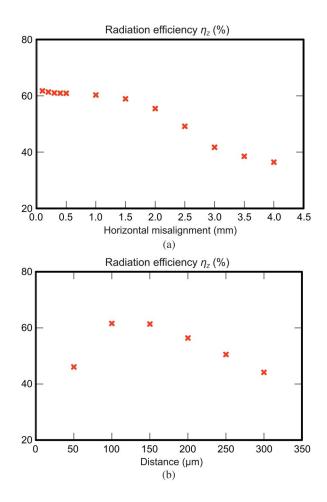
Fig. 11. *Variation of the radiation efficiency evaluated versus (a) horizontal and (b) vertical misalignment between the antenna and the circuit windings. This test allows the robustness of the antenna performances against mounting uncertainties inherent to textile electronics to be evaluated at glance. After [87].*

analog-to-digital converter (ADC), etc.]. The main advantage of such an approach is the digital modulation of the transmitted signal, and thus the flexibility of the related data treatment, as shown in [89] and [90]. The production costs of chip-based tags are mostly associated with the heterogeneous integration of the silicon chip with the antenna [85], [91], the latter being manufactured typically on a flexible substrate or a textile substrate [92]–[97].

In order to reduce the aforementioned costs and to save as much energy as possible, solutions requiring minimum amount of electronics must be pursued. For this purpose, the chipless tag family was introduced and applied to wireless sensing in the recent years [98]–[103]. It is worth recalling here that the introduced buzzword "chipless" stands for a family of tags that are actually able to implement their functionalities by using only a few lumped components and passive distributed elements, without requiring any electronic integrated circuit (IC), the chip indeed, to be mounted.

The standard chipless tags (i.e., only coding and no sensing) exploit, basically, two mechanisms, namely, time-domain and frequency-domain scattering [64]. In the first case, the elapsed time between multiple reflections is used for coding. These reflections are obtained, at tag level, by connecting the antenna to a transmission line structure with several discontinuities. In the second case, instead, the tag has a specific frequency signature that is decoded at the reader level; to implement such a behavior, a coded series of resonators is realized on the tag antenna in order to reflect or not the corresponding frequency. Both approaches (time and frequency domain) lead to tag circuits that are not necessarily easy to miniaturize; nevertheless, it is worth noting that they are compliant with the LAE paradigm.

When dynamic sensor information has to be added to the static identification, we talk about chipless RFID sensor tags; these exploit an antenna, the electrical properties of which are controlled by the change of the physical parameter to be detected. There are, mostly, two proposed approaches for this. The first one is to induce a permanent change in the antenna property when a certain critical threshold (acceleration, temperature, fluid level, etc.) is exceeded [104]; this is useful for alarm type operations. The second one exploits a sensing load, i.e., a load the impedance of which is controlled by the sensed variable, connected to an antenna [105]. In both cases, the whole wireless sensor system (tag and reader) needs to have absolute accuracy, thus limiting the system performance with respect to both distance and fabrication tolerances.

A different method has been recently introduced in [106]. In this paper, a novel sensing principle is associated to the generation of an intermodulation signal from a tag, the latter being illuminated by two waves at different frequencies. The advantage of this idea is that the tag response is generated at a known frequency, thus the presence or the absence of such a signal can hardly be misinterpreted. Similar techniques have been used in harmonic radar systems [107] and in one-bit frequency doubling tags [108].

In line with the above ideas, a novel and completely original approach was proposed in [20] and [109] to solve the issue of absolute accuracy of the most of the passive chipless RFID sensors. The chipless tag described in [109] is based on the harmonic radar concept [110], [111], i.e., on a tag that, being illuminated by a carrier at frequency f_0, is capable of generating the second harmonic $2f_0$.

A simplified block diagram of the tag is illustrated in Fig. 12. For this purpose, the sensor information is encoded as the phase difference between two signals, one acting as the reference signal for the other one. First, the tag receives a carrier at frequency f_0. Then, two equal signals at frequency $2f_0$ are generated by means of a diode-based frequency doubler and a power divider. At this point, one of the two signals is phase-shifted using a passive sensing element. Finally, the $2f_0$ signals are reirradiated by
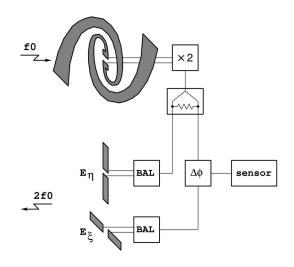
Fig. 12. *Simplified block diagram of a harmonic chipless tag with sensing capabilities. After [109].*

exploiting two orthogonally polarized antennas. With this approach, the sensor information can be extracted by a suitable reader equipped with two complex (I/Q) receivers.

It must be observed that chipless tags, due to their inherent simplicity and extremely low number of lumped components, are easily realizable on flexible substrates by means of metal laminate [29], [30] and inkjet printing technologies. In particular, antennas [112], diodes [113], and passive sensing elements [114]–[117], i.e., all the main tag components, have already been experimented in cellulose-based materials. This will soon make SSs, equipped with sensing chipless tags, feasible at affordable prices.

VI. QUASI-OPTICS

"Quasi-optics concerns the propagation of EM radiation when the size of the wavelength is comparable to the size of the optical components (e.g., lenses, mirrors, and apertures) and hence diffraction effects become significant" [118].

In optics, operations are usually performed by using lenses. At EM frequencies, by following the quasi-optics analogy, lenses become large arrays, several wavelengths in size, of unconnected elements, each of them being able to get part of the signal incident on the array (usually focused by using a Gaussian beam), to carry out an operation and to irradiate the modified signal. The sum of the radiated signals modified by the array elements ends up in a radiated beacon elaborated with respect to the incident one [118]–[120]. From a historical point of view, quasi-optics was first introduced experimentally by Heinrich Hertz in the late 1880s [120] when he demonstrated the possibility of collimating EM signals on a multiwavelength surface by using cylindrical reflectors and studying effects that, until that time, were observed only at infrared and visible EM spectrum.
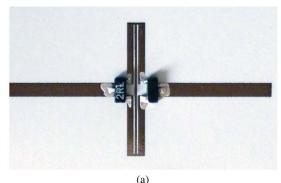
After more than 70 years, owing to the increasing demand for bandwidth and to the development of optical wavelength systems, quasi-optics experienced its palingenesis at the millimeter and submillimeter wavelengths.

We can classify the quasi-optical components into three main categories: frequency independent, frequency selective, and active devices. Frequency-independent surfaces include: delay lines, polarizing grids, hybrid junctions [121], attenuators [122], power dividers [123] and combiners [124], nonreciprocal devices [125], [126], absorbers, and calibration loads. Frequency-selective surfaces [127]–[129] include: inductive grids, capacitive grids, resonant grids, thick structures (perforated plates), and interferometers. Active devices include: oscillators [129], [130], amplifiers [131], [132], mixers [133], [134], phase shifters [135], [136], multipliers [108], [137]–[139], and switches [140]. This approach has been used so far mainly at millimeter and submillimeter frequencies because, being the size of the apparatuses proportional to the wavelength of the radiated wave, lower frequencies would have been prohibitive due to the very large size required. With the present development of technologies for LAE, however, quasi-optics has become a feasible approach even at RF and microwave bands, where large arrays of operational elements can be conceived.

Inheriting the quasi-optical approach, contact-less EM wave processing (filtering, frequency conversion, selective shielding, etc.) can be implemented even at low frequencies (in the order of megahertz). Two pioneer examples are switches [140] and the cross-dipole frequency doubler, proposed in [108], and implemented in [141], where, among various operating principles, the generation of harmonics is chosen to demonstrate the feasibility of such a component at microwave frequencies.

The layout of the proposed quasi-optical frequency doubler is shown in Fig. 13. The structure is inkjet printed on a cellulose-based substrate (a piece of photografic paper from Kodak). It consists of two crossed $\lambda/2$ dipoles. The longest dipole receives the incoming power at the fundamental frequency $f_0 = 3.5$ GHz, while the shortest one transmits the generated power at the doubled frequency $2f_0$ in an orthogonally polarized orientation. The length of the dipole operating at f_0 is 32 mm, while the $2f_0$ dipole is 16 mm long. The width of the tracks used to implement the dipoles is about 2 mm. The multiplication is provided by four diodes in a bridge configuration, thus forming a fully balanced multiplier unit.

Due to the omnidirectional nature of the dipole antenna, the harmonic signal is irradiated with the same intensity in the azimuthal plane (small differences are due to both the dielectric substrate and the planar nature of the dipole conductors). This means that the harmonic signal is reflected toward the reader and the transmitter toward other directions at the same time. A measurement of the received power at $2f_0 = 7$ GHz is reported in Fig. 13(b). Here the interrogation distance is 10 cm and the

(a)

(b)
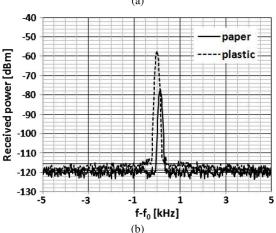
Fig. 13. *Cellulose-based prototype of the (a) crossed-dipole frequency doubling tag and (b) measured second-harmonic response. The measurements have been carried out interrogating the tag at $f_0 = 3.5$ GHz and receiving the $2f_0 = 7$ GHz frequency component. The interrogation distance is 10 cm, and the transmitted power is equal to 20 dBm. The transmitter is equipped with a two-element Yagi antenna at f_0, while the receiver uses a helix antenna at $2f_0$. The fundamental frequency dipole is 32 mm long. After [141].*

Fig. 14. *Example of quasi-optical frequency doubler based on an array of crossed dipoles. This structure is implemented on a Rogers substrate, but this can be substituted with a cellulose-based material to reduce costs and obtain mechanical flexibility.*

transmitted power is equal to 20 dBm. The transmitter (reader side) is equipped with a two-element Yagi antenna at f_0, while the receiver (reader side) uses a helix antenna at $2f_0$. In these conditions, the power received from the cellulose-based (paper) tag prototype at $2f_0$ is about −80 dBm. The same structure implemented on a plastic substrate produces more power, mainly because of a better frequency tuning. An improved efficiency could be achieved by: 1) adopting a frequency multiplier with better harmonic terminations; and 2) using directive antennas in order to address the power only toward the reader.

In the frame of the IoT, the crossed-dipole tag is a very simple one-bit tag that can alarm a system when it is placed within the range of the reader. However, with very simple modifications, this idea can be used to implement a variety of on/off sensors (the paper substrate can be easily torn when a certain mechanical strain is exceeded).

The crossed-dipole structure can also be used as a building block to form arrays, an example being shown in Fig. 14. These arrays are mainly intended for LAE applica-

tions where a complete surface can be interrogated and answered by generating a harmonic signal. The fabricated prototype uses a Rogers substrate, thus it is not flexible. However, adopting a cellulose-based substrate, a flexible structure can be built. The bending capability of cellulose-based circuits is quite good and only limited by the discrete components mounted on them.

The second example consists of a paper-based contact-less frequency doubler for harmonic RFID applications [142]. The doubler, realized on paper substrate, generates the harmonic signal by means of a single Schottky diode. The system operates at 7.5 and 15 MHz, and these frequencies are chosen, without lack of generality, to accomplish the realization of a fully organic frequency doubler exploiting paper printed coils and organic diode (pentacene-based), the present frequency limit of which is around 15 MHz [143]. Fig. 15 shows a picture of the organic tag, while Fig. 16 shows the doubling efficiency of the organic tag versus transmitting/receiving (TX/RX) distance, assuming TX power as a parameter (for the sake of completeness, it is worth noting that TX and RX antennas of the reader are equal to the RX and TX antennas of the tag, respectively).

As a final remark, it is worth noting that, to the authors' knowledge, only planar developments of quasi-optical arrays have been proposed so far; reasonably, combining quasi-optics with LAE related technologies can lead, in the near future, to new SS solutions only in a limited planar extent.

VII. SMART FLOOR

The new concept of smart cities is generating a group of new scenarios that will impose a new way people interact with the environment. This immersion brings the IoT to a new level of relationship with people and environment,
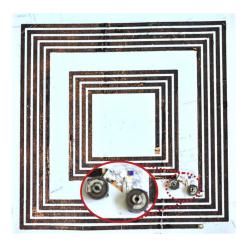
Fig. 15. *Fully organic tag prototype of the frequency doubling tag operating at 7.5–15 MHz. After [142].*

where the scenario around is aware of people's needs and can interact with their way of living.

This new concept brings IoT to areas that cover broad topics, including transportation, energy distribution, data communications, and all technologies in a sense that they will become transparent to the normal users and citizens.

The growth of smart cities will significantly increase the quality of life, the reduction of energy waste, and the availability of information by using ICT. All these developments have, as a motivation, the overall inclusion of people into the city in a way that ICT should become completely embedded into our surroundings.

These new concepts are viable only if the smart environments are enabled by IoT. IoT is actually taking shape and is growing everyday through the number of devices that get connected to each other.

Nevertheless, most of these devices that team up with each other to build up the IoT world are powered by some
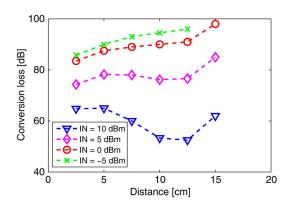
Fig. 16. *Conversion (doubling) loss of the reader tag harmonic RFID system as a function of the reader tag distance, assuming the output reader power as a parameter. After [142].*

kind of power sources. This is actually the main limitation of present IoT solutions. In order to be able to create real IoT environments, RFID, wireless power transfer (WPT), and EH devices appear as a major enabling technology for the IoT, due to their inherent simplicity and ability to provide sensing to remote objects without the need for constant powering.

In this scenario, the concept of a smart floor becomes real. Smart floors should be capable of interacting with the environment seamlessly and be aware of persons and objects on top of them.

The focus of this section is going to be smart floors as an information system, where RFID capabilities are used to provide an identification and information source embedded into the floor. This massive RFID immersion will allow to create low-power high accuracy location, navigation, and, generally speaking, an information system that can team up with passers in a noninvasive way.

In this approach, the system uses passive RFID tags that are spread beneath a flooring in order to create a map that can be read with a mobile unit that might be self-powered through energy harvested from the movements of the subject to be located [144], [145].

This new location system based on smart floors can actually be extremely competitive when compared with existing solutions; among them, we can take into consideration those based on image processing [146]; they provide a high level of information at the price of high cost, power consumption, and intrusiveness. Another approach is based on pressure sensors [146], [147], providing actually a different type of smart floors; pressure sensors are nonintrusive and can be less expensive in terms of equipment when compared to the imaging systems; however, they can still be quite expensive to install into the floors; moreover, they do not allow for seamless identification of an individual in a given space. Eventually, ultrasounds [148] and RF localization system detection [149] have to be listed; although very flexible, they require dedicated communication infrastructures, additional complexity and costs, and they are mostly sensitive to environmental changes, both in terms of accuracy and functionalities.

The just mentioned smart-floor-based system tackles most of these challenges as it is very low cost, able to have a good, and precisely predictable, accuracy, and able to detect different users in a given space.

A. RFID-Based Smart Floor Including Material Aspects

Smart environments and floors should be completely seamless to the users. This entails the need to embed electronic devices into the environment in a noninvasive way. Nevertheless, this embedding of electronic devices, mainly wireless transceivers into the floor, is not a simple matter. Floors, in fact, have been made by almost the same kind of materials for centuries; tile industries are weakly available to introduce new materials to match
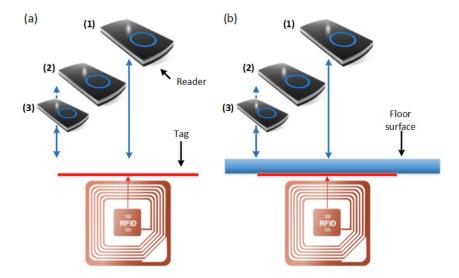
Fig. 17. *Readable distance test scheme (a) without and (b) with tile. After [144].*

requirements posed by embedded electronics. Similar considerations can be made about the consolidated techniques to build apartments and houses: they are quite mature and whatever adaptation of them to embedded electronic systems is unrealistic. The consequence of this is that constitutive elements of SSs embedded into the floors must be suitable to materials and technologies already present: materials available are those adopted for tiles and SSs must be adapted to building constraints.

In this section, we focus on the most important approaches for the characterization of the EM properties of the materials, commonly adopted in floors, and their interactions with our electronic radio systems. As an example, ceramic tiles are going to be selected as the floor element where electronic tags have to be embedded.

Ceramic tiles have been used for ages to pave floors. These make up a great deal of most floor types in public buildings, from airports to banks, therefore, they are the most interesting flooring elements to investigate.

As the first approach, the study of the influence of some materials for floor surfaces, such as ceramic, cork laminates, and wood, in the communication between the transmitter and receiver devices has to be made; in this case, low-frequency (13 MHz) inductors are going to be used as the wireless interface.

In order to do the measurements, traditional RFIDs are used, which simplify the measurement approach significantly. A test setup is shown in Fig. 17, and experimental results are summarized in Table 3. What can be seen from the results presented in the table is that none of the tested materials affect the electrical characteristics of the RFID propagation significantly.

Another approach to the measurement is to evaluate the impact of the presence of metal inclusions beneath the ground; these can appear from pipes or from buildings supporting structures. In order to test this scenario, the presence of a metal sheet close to the coil was tested to verify its impact on the RFID reading distance. The test setup is presented in Fig. 18. In this case, when the reader is aligned (1) with the center of the coil beneath the tile, a minimum distance of 10 mm between the metallic sheet and the tag must be guaranteed in order to allow for the proper reading of the tag. In the other approach, the edge aligned scenario (2), the minimum distance between the metallic sheet and the tag increases to 30 mm to allow for communication.

Table 3 Results From Average Readable Distance Test

Tile type	Aligned (1)	Edge aligned (2)	Unaligned (3)
Without tile	4.5 cm	4.3 cm	2.4cm
Ceramic tile 4 mm thick	4.5 cm	4.3 cm	2.4cm
Ceramic tile 6 mm thick	4.5 cm	4.3 cm	2.4cm
Ceramic tile 7 mm thick	4.5 cm	4.3 cm	2.4cm
Cork laminate 5 mm thick	4.5 cm	4.3 cm	2.2cm
Wooden floor 8 mm thick	4.3 cm	4.2 cm	2.2cm
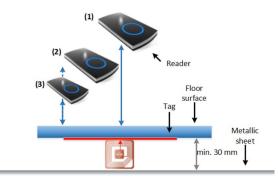
Fig. 18. *Readable distance test scheme with metallic sheet presence. After [144].*

This test allows us to conclude that, in a real scenario, when considering the typical distance from the antenna to the floor, where the reader is usually embedded into a mobile device, a minimum distance from the tile to any metal inclusions in the ground has to be guaranteed. As seen, this safe distance should be at least 30 mm.

B. Navigation Approach Based on Smart Floors

A meaningful example of the applications of the smart floor concept is the demonstrator developed at the University of Aveiro, Aveiro, Portugal. It allows a nomadic RFID reader on top of it to be identified and localized by means of a smart floor.

The prototype in the present state is composed of four main components: a set of tagged tiles forming the smart floor, the nomadic reader, a wireless communication device to connect the reader to the information system, and a computer on which the location and navigation software is installed. A potential scenario of application for the monitoring of elderly housing is shown in Fig. 19. In the demonstrator presented, the RFID reader is connected to the radio unit and inserted into the user's shoes in order to read the passive tags as the user walks around a given area. The tile IDs are detected by the reader and sent to a

Fig. 19. *Example of smart floor application scenario.*

centralized information system using, without lack of generality, a nonstandard wireless communication system, purposely developed in the laboratory.

The software in the computer then decodes the ID of the smart tile, and correlates it with a preloaded referenced map, thus being able to provide the precise position of the walker in the environment.

The location can be subsequently translated to a navigation system, and combined with artificial intelligence to allow navigation, situation awareness, etc.
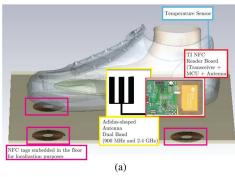
Plenty of applications and related solutions can be envisioned on the basis of this smart floor approach: robot control, automatic pallet managements, prevention in harsh environments, and so forth.

VIII. SMART SHOES AND BAN

The aforementioned smart floor (Section VII), initially proposed to localize moving subjects in a given area, actually allows for the implementation of a very low-cost, accurate, and reliable system able to provide each kind of detectable information from an equipped area to the Internet.

To exploit the potentiality of a smart floor, it is obvious that developing purposely conceived communication units acting, on the one hand, as readers for tag tiles and, on the other hand, as communication systems with the web is needed.

One of the investigated systems for this is the so-called smart shoe [21]. Smart shoes are intended as shoes incorporating a reader able to illuminate the tag embedded into the tile and a wireless interface to communicate data to the web. To this extent, smart shoes can be seen as hubs of BANs able, on the one hand, to get information from the surrounding environment (body included) and, on the other hand, to communicate the collected information to the Internet. According to common buzzwords, they can be seen as the second technological layer, just above the first physical one consisting of tagged objects, in a hierarchical picture of IoT. Alternatively, they provide the cloud of nomadic readers with networking capabilities required for the implementation of the so-called networked RFID (N–RFID) [150] strategies. In the referred structure [21], the necessary antenna system is inkjet printed on paper, thus exploiting the process described in Section II, and it is a trademark-logo-shaped dual-band antenna working at 900 MHz and 2.4 GHz. A block diagram and its equivalent realization are shown in Fig. 20: it is worth noting that the reader itself works at 13.5 MHz exploiting the near-field communication (NFC) concept, according to what is stated in the smart floor section; as a consequence, there are no issues about the interference between the dual-band antenna and the NFC reader. Moreover, an approach similar to [145] is proposed to power the reader: the idea here is as well to harvest the energy produced by the human walking by means of piezoelectric materials embedded into the sole of the shoe.

(a)

(b)

(c)
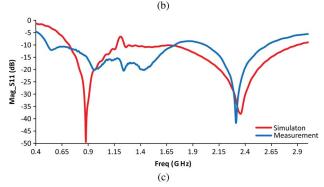
Fig. 20. *(a) Block diagram, (b) photo, and (c) dual-band antenna results of the smart shoe that works as an energy autonomous reader for smart floors able to collect the localization data from the smart tiles and send them to a centralized system. After [21].*

(a)

(b)

Fig. 21. *(a) Mounted electronics on shoe with piezoelectric energy harvester and block diagram of the transmitting tag, driven by the EH unit. (b) Photograph of the assembled prototype showing the key components packaged on an organic flexible substrate. After [145].*

Fig. 21(a) shows the shoe with the mounted electronics consisting of a trademark-logo antenna inkjet printed on a flexible paper substrate together with the circuits in which the components [Fig. 21(b)] are fixed by conductive epoxy glue. A piezoelectric push button was embedded in the shoe to scavenge energy from human walking strike. This energy is then collected by a simple electronic circuit based on off-the-shelf components and used to supply the RFID transmitter. As reported in [145], the system is composed by the power-generator/energy-conversion device, the energy storage device, the power regulator circuit, and the RF transmitter that can broadcast the tag ID and the stored information.

When the push button is pressed, an inner spring is compressed and when the pressure exceeds a fixed threshold, the spring-loaded hammer will be released to deliver the dynamic mechanical force to the piezoelectric compo-

nent; once the hammer strikes the piezoelectric element, a pressure wave is generated and reflected a few times between the hammer and the element, creating a mechanical resonance [151]. As a result, the generated output voltage follows closely an alternating current (ac) signal course. The signal out of the piezoelectric element is characterized by high voltage and low current; therefore, since this RF circuitry requires lower voltages at higher currents, a step-down transformer is used for better impedance matching to the following circuitry.

An energy storage device is adopted to collect the electrical energy and to provide it to the transmitter even when the external power source is temporarily unavailable; in fact, the RF transmitter takes tens of milliseconds to transmit one complete word, while the piezoelectric push button harvests energy in a transient time. To store the collected energy, a 4.7-μF tank capacitor was chosen. An ac–dc full wave diode bridge rectifier is used to convert the ac voltage from a piezoelectric element into a dc source and a dc linear regulator is added to adapt the dc voltage across the capacitor to 3-V transmitter operating voltage. For this purpose, a MAX666 low-dropout linear regulator, providing a stable 3-V supply until the tank capacitor charge is drained below the reference level of the dc linear regulator itself, is chosen. The transmitter used, the LINX

Table 4 Shoe Energy Performance Summary

Energy provided by the pushbutton	stored in the capacitor	848.4 μJ
Utilized energy	below 2.7 V capacitor voltage, the active RFID tag stops transmitting	17.1 μJ
Available energy	848.4 μJ − 17.1 μJ	831.3 μJ
Energy required by the circuit for a one-word transmission	Power Needed for 50 ms operation: 9 mW	450 μJ

TXE-433-KH2 [152], is able to combine a highly optimized RF transmitter with an onboard encoder.

The maximum voltage across the capacitor is 19 V; therefore, the energy stored in the capacitor can be calculated to be close to 850 μJ. After the strike, the regulator becomes active, providing power to the tag circuitry; its 3-V output remains constant before it starts to drop to zero according to the capacitor discharge rate. When the voltage across the capacitor drops below 2.7 V, the active tag stops transmitting. The resulting energy left in the capacitor and not used is about 17 μJ. The required time for successful completion has been measured to be close to 50 ms; the total circuit energy consumption for a 50-ms transmission is then approximately 9 mW (Section IV). Consequently, the energy required for a one word transmission is approximately 450 μJ, much less than the power collected by the energy conversion and regulation unit. The energy performance of the smart shoe is summarized in Table 4.

"Smart" solutions like smart shoes are suitable for smart floor interaction and for collecting information (position included) from them. The smart shoe has actually a double functionality: on the one hand, it is a means to collect information from the environment (smart floor included) and from the on-body sensors and apparatuses; on the other hand, it transfers data to the Internet, via WiFi, for instance, while likely receiving control and feedback signals. To this extent, it contributes to the development of wearable electronics for body-centric communication systems, according to the scheme of Fig. 22. This class of devices is based on the BAN concept (see Fig. 22): a network of wearable computational units, sensors, energy harvesters, and transceivers networked and implemented around the human body. The information is acquired by means of several sensors and, after processing and digital modulation, is transmitted by means of off-body transceivers. The most attractive application of BANs and wearable systems, so far, is probably health monitoring [88], but the number of foreseeable applications is constantly increasing (personal fitness monitoring, personal audio system, personal alarm set, and so on). Today, a big contribution to the BANs is being given by the interaction with portable devices such as smartphones which can be exploited as gateways.

Low-power ICs, sensors, and wireless communications are key elements of the BANs, and the use of EH techniques

and mixed technologies such as RFID, as in the above described smart shoe, can contribute to the development of autonomous BANs. According to the notation adopted in [74] and [153], BANs can exploit wireless communication not only for off-body communications with other systems, but also for on-body and in-body communications to interconnect networked devices [wireless body area networks (WBANs)]. Generally, the proposed WBAN solutions are based on existing communication standards; but, recently, the IEEE 802.15.6 working group has provided a standard for short-range, wireless communications in the vicinity of, or inside, a human body based on ISM bands for BANs [154]. In this rapidly evolving scenario, smart shoes can be seen, in turn, as a promising mean to interface information coming from smart floors to the BANs and ultimately to the Internet, directly or via portable devices.

IX. ENERGY SKIN

Within SSs, another area is emerging, which requires dc energy. The availability of devices that can actually increase the battery life of our electronic gadgets [155] is the holy grail of IoT; the idea is actually to convert energy from

Fig. 22. *BAN integrated with IoT: the data sensed on the human body and collected by smart devices are sent to the Internet through the standard protocols like third generation/fourth generation (3G/4G) and WiFi.*
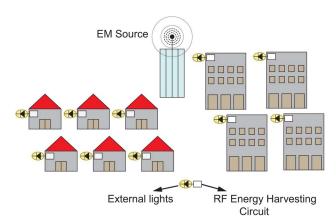
Fig. 23. *EM source, likely supplied by solar energy in order to keep eco-friendliness, is used to broadcast energy to the urban neighborhood where houses are equipped with local RF–dc converters able to power supply local devices (external lamps in the example).*

different sources, as solar to dc [156], EM radio signals to dc, thermal to dc [156], vibration to dc [157], etc. [158].

Portable converters from solar to dc are today commonly available on the market at very low prize. Similar portable scavengers can be easily envisioned even for other energy converters. On this basis, plenty of sophisticated and more complex solutions are being proposed. Just as an example, we report the solution in Fig. 23, where the solar energy is used to supply an EM source that broadcasts RF energy in the surrounding urban environment; this energy is converted to dc energy at building level and, in turn, exploited to supply external local lamps. This approach was conceived for space applications [159], [160], but could be used in certain conditions in earth stations as well. The concept can actually move forward and be integrated into a smart skin, in the sense that all devices could be powered up by the skin energy conversion that a person can carry

(vibrational, thermal, RF, etc.). This will become possible by integrating paper and textile radio electronic circuits in the smart skin surfaces of our clothes.

X. CONCLUSION

This paper focused on smart surfaces concept. This concept was introduced as a way to provide solutions to societal instances by means of technologies and techniques that today are gaining great momentum thanks to IoT development. The foreseeable IoT evolution, in fact, calls for systems and subsystems that are more and more energy conservative, fully autonomous (battery-less and not connected to the grid), as recyclable as possible (ultimately 100% biodegradable), flexible, and extremely low cost to be embedded in as many objects as possible. In order to meet these expectations, several technologies have to be implemented concurrently. This paper illustrates some of them, namely, recyclable and reliable materials, printing and R2R compatible techniques, conformal antennas on unconventional materials, and chipless RFID.

Subsequently, it has been shown how these technologies enable a fairly new branch of IoT development, represented just by SSs.

SSs are in their infancy; nevertheless, some examples, enabled by the aforementioned technologies, can be cited. Among them, the way to inherit "quasi-optics" and extend them to RF frequencies even below gigahertz is shown, and the smart floor implementation as a means to provide precise localization as well as assisted paths platforms has been reported. Energy skin as a means to convert several energy sources (solar, thermal, vibrational, EM, etc.) in electric one is described as well.

The reported examples must be researched as a new paradigm for potential promising developments aiming at providing new LAE solutions for the future networked society. ∎

REFERENCES

[1] European Union (EU), "The ICT sector and RDI," Digital Agenda Scoreboard, 2012. [Online]. Available: http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/scoreboard_ict_sector_and_rdi.pdf

[2] J. Manyika *et al.*, "How the Internet of Things is more like the Industrial Revolution than the Digital Revolution," MGI Rep., 2013. [Online]. Available: http://www.mckinsey.com/insights/business_technology/disruptive_technologies

[3] D. Evans, "The Internet of Things: How the next evolution of the internet is changing everything," CISCO White papers, 2011. [Online]. Available: http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

[4] I. G. Smith, "The Internet of Things 2012 New Horizons," IERC European Research Cluster on the Internet of Things, 2012.

[5] J. Morse, "Nanofabrication technologies for roll-to-roll processing," in *Proc. NIST-NNN Workshop*, Boston, MA, USA, Sep. 2011, pp. 1–31.

[6] A. Arias, J. MacKenzie, I. McCulloch, J. Rivnay, and A. Salleo, "Materials and applications for large area electronics: Solution-based approaches," *Chem. Rev.*, vol. 110, no. 1, pp. 3–24, Jan. 2010.

[7] T. Kampke, B. Kluge, and M. Strobel, *Field and Service Robotics*. New York, NY, USA: Springer-Verlag, 2008, ser. Springer Tracts in Advanced Robotics.

[8] M. Shur, "Giant area and flexible electronics," in *Proc. IEEE Nanotechnol. Mater. Devices Conf.*, Oct. 2006, vol. 1, pp. 264–265.

[9] M. Steer *et al.*, "Global modeling of spatially distributed microwave and millimeter-wave systems," *IEEE Trans. Microw. Theory Tech.*, vol. 47, no. 6, pp. 830–838, Jun. 1999.

[10] A. Rida, L. Yang, R. Vyas, and M. Tentzeris, "Conductive inkjet-printed antennas on flexible low-cost paper-based substrates for RFID and WSN applications," *IEEE Antennas Propag. Mag.*, vol. 51, no. 3, pp. 13–23, Jun. 2009.

[11] F. Alimenti, P. Mezzanotte, M. Dionigi, M. Virili, and L. Roselli, "Microwave circuits in paper substrates exploiting conductive adhesive tapes," *IEEE Microw. Wireless Compon. Lett.*, vol. 22, no. 12, pp. 660–662, Dec. 2012.

[12] D. Tobjörk and R. Österbackaa, "Paper electronics," *Adv. Mater.*, vol. 23, no. 17, pp. 1935–1961, 2011.

[13] P. Calvert, "Inkjet printing for materials and devices," *Chem. Mater.*, vol. 13, no. 10, pp. 3299–3305, 2001.

[14] G. Orecchini *et al.*, "Inkjet printed organic transistors for sustainable electronics," in *Proc. 60th Electron. Compon. Technol. Conf.*, Jun. 2010, pp. 985–989.

[15] L. Yang, G. Orecchini, G. Shaker, H.-S. Lee, and M. Tentzeris, "Battery-free RFID-enabled wireless sensors," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2010, pp. 1528–1531.

[16] G. Shaker, S. Safavi-Naeini, N. Sangary, and M. Tentzeris, "Inkjet printing of ultrawideband (UWB) antennas on paper-based substrates," *IEEE Antennas*

Wireless Propag. Lett., vol. 10, pp. 111–114, 2011.

[17] R. Vyas *et al.*, "Inkjet printed, self powered, wireless sensors for environmental, gas, and authentication-based sensing," *IEEE Sensors J.*, vol. 11, no. 12, pp. 3139–3152, Dec. 2011.

[18] G. Orecchini *et al.*, "Design and fabrication of ultra-low cost radio frequency identification antennas and tags exploiting paper substrates and inkjet printing technology," *IET Microw. Antennas Propag.*, vol. 5, no. 8, pp. 993–1001, Jun. 2011.

[19] S. Kim *et al.*, "Monopole antenna with inkjet-printed EBG array on paper substrate for wearable applications," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 663–666, 2012.

[20] S. Kim *et al.*, "No battery required: Perpetual RFID-enabled wireless sensors for cognitive intelligence applications," *IEEE Microw. Mag.*, vol. 14, no. 5, pp. 66–77, Jul. 2013.

[21] C. Mariotti, V. Lakafosis, M. Tentzeris, and L. Roselli, "An IPv6-enabled wireless shoe-mounted platform for health-monitoring," in *Proc. IEEE Top. Conf. Wireless Sensors Sensor Netw.*, Jan. 2013, pp. 46–48.

[22] K. Sangkil *et al.*, "Inkjet-printed antennas, sensors and circuits on paper substrate," *IET Microw. Antennas Propag.*, vol. 7, no. 10, pp. 858–868, 2013.

[23] B. Cook, J. Cooper, and M. Tentzeris, "Multi-layer RF capacitors on flexible substrates utilizing inkjet printed dielectric polymers," *IEEE Microw. Wireless Compon. Lett.*, vol. 23, no. 7, pp. 353–355, Jul. 2013.

[24] B. Cook, J. Cooper, S. Kim, and M. Tentzeris, "A novel inkjet-printed passive microfluidic RFID-based sensing platform," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2013, DOI: 10.1109/MWSYM.2013.6697592.

[25] B. Cook, B. Tehrani, J. Cooper, and M. Tentzeris, "Multilayer inkjet printing of millimeter-wave proximity-fed patch arrays on flexible substrates," *IEEE Antennas Wireless Propag. Lett.*, vol. 12, pp. 1351–1354, 2013.

[26] C. Mariotti, B. Cook, L. Roselli, and M. Tentzeris, "High-performance inkjet-printed metal-insulator-metal (MIM) capacitors on lossy silicon substrate *IEEE Microw. Wireless Compon. Lett.*, 2014.

[27] D. Angmo, T. T. Larsen-Olsen, M. Jorgensen, R. R. Saszndergaard, and F. C. Krebs, "Roll-to-roll inkjet printing and photonic sintering of electrodes for ITO free polymer solar cell modules and facile product integration," *Adv. Energy Mater.*, vol. 3, no. 2, pp. 172–175, 2013.

[28] R. Saszndergaard, M. Hessel, D. Angmo, T. T. Larsen-Olsen, and F. C. Krebs, "Roll-to-roll fabrication of polymer solar cells," *Mater. Today*, vol. 15, no. 1–2, pp. 36–49, 2012.

[29] F. Alimenti *et al.*, "A 1.2 v, 0.9 mW UHF VCO based on hairpin resonator in paper substrate and Cu adhesive tape," *IEEE Microw. Wireless Compon. Lett.*, vol. 23, no. 4, pp. 214–216, Apr. 2013.

[30] F. Alimenti *et al.*, "24 GHz single-balanced diode mixer exploiting cellulose-based materials," *IEEE Microw. Wireless Compon. Lett.*, vol. 23, no. 11, pp. 596–598, Nov. 2013.

[31] C. Mariotti *et al.*, "Modeling and characterization of copper tape microstrips on paper substrate and application to 24 GHz branch-line couplers," in *Proc. Eur. Microw. Conf.*, Oct. 2013, pp. 794–797.

[32] L. Guoyu and W. Zhenkai, "Summarize of rfid technology and typical application," in

Proc. Cross Strait Quad-Regional Radio Sci. Wireless Technol. Conf., Jul. 2011, vol. 2, pp. 1032–1036.

[33] L. Wang, L. D. Xu, Z. Bi, and Y. Xu, "Data cleaning for RFID and WSN integration," *IEEE Trans. Ind. Inf.*, vol. 10, no. 1, pp. 408–418, Feb. 2014.

[34] H. Deng *et al.*, "Design of sensor-embedded radio frequency identification (SE-RFID) systems," in *Proc. IEEE Int. Conf. Mechatron. Autom.*, Jun. 2006, pp. 792–796.

[35] S. Mohamad, F. Tang, A. Amira, A. Bermak, and M. Benammar, "A low power oscillator based temperature sensor for RFID applications," in *Proc. 5th Asia Symp. Quality Electron. Design*, Aug. 2013, pp. 50–54.

[36] D. Hotte, R. Siragusa, Y. Duroc, and S. Tedjini, "Performance optimization for the choice of impedances in the design of passive RFID tags," in *Proc. IEEE-APS Top. Conf. Antennas Propag. Wireless Commun.*, Sep. 2013, pp. 1349–1352.

[37] M. Donelli, "Guidelines for the design and optimization of wireless sensors based on the modulated scattering technique," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 7, pp. 1824–1833, Jul. 2014.

[38] A. Ricci, M. Grisanti, I. De Munari, and P. Ciampolini, "Improved pervasive sensing with RFID: An ultra-low power baseband processor for UHF tags," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 12, pp. 1719–1729, Dec. 2009.

[39] L. Chen *et al.*, "Range extension of passive wake-up radio systems through energy harvesting," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2013, pp. 1549–1554.

[40] S. Sauer, A. Turke, A. Weder, and W.-J. Fischer, "Inkjet printed humidity threshold monitoring sensor solution with irreversible resistance change for passive RFID applications," in *Proc. IEEE Sensors*, Nov. 2013, DOI: 10.1109/ICSENS.2013.6688307.

[41] N. Cho, S.-J. Song, S. Kim, S. Kim, and H.-J. Yoo, "A 5.1-uW UHF RFID tag chip integrated with sensors for wireless environmental monitoring," in *Proc. 31st Eur. Solid-State Circuits Conf.*, Sep. 2005, pp. 279–282.

[42] M. Philipose *et al.*, "Battery-free wireless identification and sensing," *IEEE Perv. Comput.*, vol. 4, no. 1, pp. 37–45, Jan. 2005.

[43] D. De Donno, L. Catarinucci, and L. Tarricone, "A battery-assisted sensor-enhanced RFID tag enabling heterogeneous wireless sensor networks," *IEEE Sensors J.*, vol. 14, no. 4, pp. 1048–1055, Apr. 2014.

[44] N. Guilar, T. Kleeburg, A. Chen, D. Yankelevich, and R. Amirtharajah, "Integrated solar energy harvesting and storage," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 5, pp. 627–637, May 2009.

[45] D. Brunelli, C. Moser, L. Thiele, and L. Benini, "Design of a solar-harvesting circuit for batteryless embedded systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 11, pp. 2519–2528, Nov. 2009.

[46] D. Kwon, G. Rincon-Mora, and E. Torres, "Harvesting ambient kinetic energy with switched-inductor converters," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 7, pp. 1551–1560, Jul. 2011.

[47] S. Mehraeen, S. Jagannathan, and K. Corzine, "Energy harvesting from vibration with alternate scavenging circuitry and tapered cantilever beam,"

IEEE Trans. Ind. Electron., vol. 57, no. 3, pp. 820–830, Mar. 2010.

[48] T. Le, K. Mayaram, and T. Fiez, "Efficient far-field radio frequency energy harvesting for passively powered sensor networks," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1287–1302, May 2008.

[49] G. Papotto, F. Carrara, and G. Palmisano, "A 90-nm CMOS threshold-compensated RF energy harvester," *IEEE J. Solid-State Circuits*, vol. 46, no. 9, pp. 1985–1997, Sep. 2011.

[50] S. Rahimizadeh, S. Korhummel, B. Kaslon, and Z. Popovic, "Scalable adaptive wireless powering of multiple electronic devices in an over-moded cavity," in *Proc. IEEE Wireless Power Transfer*, May 2013, pp. 84–87.

[51] S. Scorcioni, L. Larcher, A. Bertacchini, L. Vincetti, and M. Maini, "An integrated RF energy harvester for UHF wireless powering applications," in *Proc. IEEE Wireless Power Transfer*, May 2013, pp. 92–95.

[52] Z. Popovic, "Far-field wireless power delivery and power management for low-power sensors," in *Proc. IEEE Wireless Power Transfer*, May 2013, DOI: 10.1109/WPT.2013.6556867.

[53] E. Falkenstein, M. Roberg, and Z. Popovic, "Low-power wireless power delivery," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 7, pp. 2277–2286, Jul. 2012.

[54] S. Roundy, P. K. Wright, and J. Rabaey, "A study of low level vibrations as a power source for wireless sensor nodes," *Comput. Commun.*, vol. 26, no. 11, pp. 1131–1144, 2003.

[55] C. D. Richards, M. J. Anderson, D. F. Bahr, and R. F. Richards, "Efficiency of energy conversion for devices containing a piezoelectric component," *J. Micromech. Microeng.*, vol. 14, no. 5, 2004, DOI: 10.1088/0960-1317/14/5/009.

[56] S. R. Anton and H. A. Sodano, "A review of power harvesting using piezoelectric materials (2003–2006)," *Smart Mater. Struct.*, vol. 16, no. 3, 2007, DOI: 10.1088/0964-1726/16/3/R01.

[57] Cymbet Corporation, "Energy harvesting and storage." [Online]. Available: http://www.cymbet.com/design-center/energy-harvesting.php#

[58] T. Armstrong, "Energy harvesting applications are everywhere," RTC, May 2011. [Online]. Available: http://www.rtcmagazine.com/articles/view/102133

[59] D. Puccinelli and M. Haenggi, "Wireless sensor networks: Applications and challenges of ubiquitous sensing," *IEEE Circuits Syst. Mag.*, vol. 5, no. 3, pp. 19–31, 2005.

[60] L. Roselli *et al.*, "WPT, RFID and energy harvesting: Concurrent technologies for the future networked society," in *Proc. Asia-Pacific Microw. Conf.*, Nov. 2013, pp. 462–464.

[61] H. Liu, M. Bolic, A. Nayak, and I. Stojmenovic, "Taxonomy and challenges of the integration of RFID and wireless sensor networks," *IEEE Network*, vol. 22, no. 6, pp. 26–35, Nov. 2008.

[62] D. Dardari, R. D'Errico, C. Roblin, A. Sibille, and M. Win, "Ultrawide bandwidth RFID: The next generation?" *Proc. IEEE*, vol. 98, no. 9, pp. 1570–1582, Sep. 2010.

[63] C. Balanis, *Antenna Theory: Analysis and Design*. New York, NY, USA: Wiley, 2005, ser. Knovel Library.

[64] K. Finkenzeller and D. Müller, *RFID Handbook: Fundamentals and Applications*

in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication. New York, NY, USA: Wiley, 2010.

[65] E. Perret, S. Tedjini, and R. Nair, "Design of antennas for UHF RFID tags," *Proc. IEEE*, vol. 100, no. 7, pp. 2330–2340, Jul. 2012.

[66] K. V. S. Rao, P. Nikitin, and S. Lam, "Antenna design for UHF RFID tags: A review and a practical application," *IEEE Trans. Antennas Propag.*, vol. 53, no. 12, pp. 3870–3876, Dec. 2005.

[67] H. Chaabane, E. Perret, and S. Tedjini, "A methodology for the design of frequency and environment robust UHF RFID tags," *IEEE Trans. Antennas Propag.*, vol. 59, no. 9, pp. 3436–3441, Sep. 2011.

[68] A. Babar, A. Elsherbeni, L. Sydanheimo, and L. Ukkonen, "RFID tags for challenging environments: Flexible high-dielectric materials and ink-jet printing technology for compact platform tolerant RFID tags," *IEEE Microw. Mag.*, vol. 14, no. 5, pp. 26–35, Jul. 2013.

[69] M. Scarpello, I. Kazani, C. Hertleer, H. Rogier, and D. Vande Ginste, "Stability and efficiency of screen-printed wearable and washable antennas," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 838–841, 2012.

[70] V. Lakafosis *et al.,* "Progress towards the first wireless sensor networks consisting of inkjet-printed, paper-based RFID-enabled sensor tags," *Proc. IEEE*, vol. 98, no. 9, pp. 1601–1609, Sep. 2010.

[71] C. Mariotti, G. Orecchini, M. Virili, F. Alimenti, and L. Roselli, "Wireless localization in buildings by smart tiles," in *Proc. IEEE Workshop Environ. Energy Struct. Monitoring Syst.*, Sep. 2012, pp. 7–11.

[72] H. J. Song *et al.,* "A method for improving the efficiency of transparent film antennas," *IEEE Antennas Wireless Propag. Lett.*, vol. 7, pp. 753–756, 2008.

[73] F. Colombel *et al.,* "Ultra-thin metal layer, ITO film and ITO/Cu/ITO multilayer towards transparent antenna," *IET Sci. Meas. Technol.*, vol. 3, no. 3, pp. 229–234, May 2009.

[74] D. Zito and D. Pepe, "System-on-a-chip radio transceivers for 60-GHz wireless body-centric communications," in *Ultra-Wideband and 60 GHz Communications.* New York, NY, USA: Springer-Verlag, 2014, pp. 177–187.

[75] L. Martin *et al.,* "Effect of permittivity and permeability of a flexible magnetic composite material on the performance and miniaturization capability of planar antennas for RFID and wearable wireless applications," *IEEE Trans. Compon. Packag. Technol.*, vol. 32, no. 4, pp. 849–858, Dec. 2009.

[76] D. Kim and J. Yeo, "Low-profile RFID tag antenna using compact AMC substrate for metallic objects," *IEEE Antennas Wireless Propag. Lett.*, vol. 7, pp. 718–720, 2008.

[77] C. Kakoyiannis and P. Constantinou, "Compact WSN antennas of analytic geometry based on Chebyshev polynomials," in *Proc. Antennas Propag. Conf.*, Nov. 2012, DOI: 10.1109/LAPC.2012.6402976.

[78] H. Visser and R. Vullers, "RF energy harvesting and transport for wireless sensor network applications: Principles and requirements," *Proc. IEEE*, vol. 101, no. 6, pp. 1410–1423, Jun. 2013.

[79] M. Pinuela, P. Mitcheson, and S. Lucyszyn, "Ambient RF energy harvesting in urban and semi-urban environments," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 7, pp. 2715–2726, Jul. 2013.

[80] K. M. Farinholt, G. Park, and C. Farrar, "RF energy transmission for a low-power wireless impedance sensor node," *IEEE Sensors J.*, vol. 9, no. 7, pp. 793–800, Jul. 2009.

[81] J. Hagerty, F. Helmbrecht, W. McCalpin, R. Zane, and Z. Popovic, "Recycling ambient microwave energy with broad-band rectenna arrays," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1014–1024, Mar. 2004.

[82] A. Collado and A. Georgiadis, "Conformal hybrid solar and electromagnetic (EM) energy harvesting rectenna," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 8, pp. 2225–2234, Aug. 2013.

[83] T. Wu, R. Li, and M. Tentzeris, "A scalable solar antenna for autonomous integrated wireless sensor nodes," *IEEE Antennas Wireless Propag. Lett.*, vol. 10, pp. 510–513, 2011.

[84] A. Sample, J. Braun, A. Parks, and J. Smith, "Photovoltaic enhanced UHF RFID tag antennas for dual purpose energy harvesting," in *Proc. IEEE Int. Con. RFID*, Apr. 2011, pp. 146–153.

[85] F. Alimenti *et al.,* "A new contactless assembly method for paper substrate antennas and UHF RFID chips," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 3, pp. 627–637, Mar. 2011.

[86] W. Pachler *et al.,* "A silver ink-jet printed UHF booster antenna on flexible substratum with magnetically coupled die on-chip antenna," in *Proc. 7th Eur. Conf. Antennas Propag.*, Apr. 2013, pp. 1730–1733.

[87] M. Virili, H. Rogier, F. Alimenti, P. Mezzanotte, and L. Roselli, "Wearable textile antenna magnetically coupled to flexible active electronic circuits," *IEEE Antennas Wireless Propag. Lett.*, vol. 13, pp. 209–212, 2014.

[88] H.-J. Yoo, "Your heart on your sleeve: Advances in textile-based electronics are weaving computers right into the clothes we wear," *IEEE Solid-State Circuits Mag.*, vol. 5, no. 1, pp. 59–70, Mar. 2013.

[89] X. Jingtian *et al.,* "Low-cost low-power UHF RFID tag with on-chip antenna," *J. Semicond.*, vol. 30, no. 7, pp. 075012/1–075012/6, Jul. 2009.

[90] M. Law, A. Bermak, and H. Luong, "A sub-μW embedded CMOS temperature sensor for RFID food monitoring application," *IEEE J. Solid-State Circuits*, vol. 45, no. 6, pp. 1246–1255, Jun. 2010.

[91] Alien Technology Corporation, "Fluidic self assembly," White Paper, 1999. [Online]. Available: http://www.alientechnology.com

[92] C. Hertleer, H. Rogier, L. Vallozzi, and L. V. Langenhove, "A textile antenna for off-body communication integrated into protective clothing for firefighters," *IEEE Trans. Antennas Propag.*, vol. 57, no. 4, pp. 919–925, Apr. 2009.

[93] X. Li, J. Liao, Y. Yuan, and D. Yu, "Eye-shaped segmented reader antenna for near-field UHF RFID applications," *Progr. Electromagn. Res.*, vol. 114, pp. 481–493, 2011.

[94] J.-J. Tiang, M. Islam, N. Misran, and J. S. Mandeep, "Circular microstrip slot antenna for dual-frequency RFID application," *Progr. Electromagn. Res.*, vol. 120, pp. 499–512, 2011.

[95] Y. Amin, Q. Chen, H. Tenhunen, and L.-R. Zheng, "Performance-optimized quadrate bowtie RFID antennas for cost-effective and eco-friendly industrial applications," *Progr. Electromagn. Res.*, vol. 126, pp. 49–64, 2012.

[96] Y. Amin, Q. Chen, L.-R. Zheng, and H. Tenhunen, "Development and analysis of flexible UHF RFID antennas for "green" electronics," *Progr. Electromagn. Res.*, vol. 130, pp. 1–15, 2012.

[97] F. Viani, M. Salucci, G. Olivieri, F. Robol, and A. Massa, "Design of UHF RFID/GPS fractal antenna for logistic management," *J. Electromagn. Waves Appl.*, vol. 26, no. 4, pp. 480–492, 2012.

[98] A. Springer, R. Weigel, A. Pohl, and F. Seifert, "Wireless identification and sensing using surface acoustic wave devices," *Mechatronics*, vol. 9, no. 7, pp. 745–756, Oct. 1999.

[99] K. Chang, Y. Kim, Y. Kim, and Y. Yoon, "Functional antenna integrated with relative humidity sensor using synthesised polymide for passive RFID sensing," *Electron. Lett.*, vol. 43, no. 5, pp. 1918–1923, May 2007.

[100] S. Shrestha, M. Balachandran, M. Agarwal, V. Phoha, and K. Varahramyan, "A chipless RFID sensor system for cyber centric monitoring applications," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 5, pp. 1303–1309, May 2009.

[101] C. Occhiuzzi, S. Cippitelli, and G. Marrocco, "Modeling, design and experimentation of wearable RFID sensor tag," *IEEE Trans. Antennas Propag.*, vol. 58, no. 8, pp. 2490–2498, Aug. 2010.

[102] A. Ramos, A. Lazaro, D. Girbau, and R. Villarino, "Time-domain measurement of time-coded UWB chipless RFID tags," *Progr. Electromagn. Res.*, vol. 116, pp. 313–331, 2011.

[103] R. Nair, E. Perret, and S. Tedjini, "Temporal multi-frequency encoding technique for chipless RFID applications," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Montreal, QC, Canada, Jun. 2012, DOI: 10.1109/MWSYM. 2012.6259483.

[104] R. Bhattacharyya, C. Floerkemeier, and S. Sarma, "Low-cost, ubiquitous RFID-tag-antenna-based sensing," *Proc. IEEE*, vol. 98, no. 9, pp. 1593–1600, Sep. 2010.

[105] R. Potyrailo *et al.,* "Development of radio-frequency identification sensors based on organic electronic sensing materials for selective detection of toxic vapors," *J. Appl. Phys.*, vol. 106, no. 12, pp. 124 902–124 902-6, Dec. 2009.

[106] V. Viikari and H. Seppa, "RFID MEMS sensor concept based on intermodulation distortion," *IEEE Sensor J.*, vol. 9, no. 12, pp. 1918–1923, Dec. 2009.

[107] J. Riley *et al.,* "Tracking bees with harmonic radar," *Nature*, vol. 379, pp. 29–30, Jan. 1996.

[108] S. Helbing *et al.,* "Design and verification of a novel crossed dipole structure for quasi-optical frequency doublers," *IEEE Microw. Guided Wave Lett.*, vol. 10, no. 3, pp. 105–107, Mar. 2000.

[109] F. Alimenti and L. Roselli, "Theory of zero-power RFID sensors based on harmonic generation and orthogonally polarized antennas," *Progr. Electromagn. Res.*, vol. 134, pp. 337–357, 2013.

[110] G. Lovei, I. Stringer, C. Devine, and M. Cartellieri, "Harmonic radar—A method using inexpensive tags to study invertebrate movement on land," *New Zealand J. Ecology*, vol. 21, no. 2, pp. 187–193, 1997.

[111] B. Colpitts and G. Boiteau, "Harmonic radar transceiver design: Miniature tags for insect tracking," *IEEE Trans. Antennas Propag.*, vol. 52, no. 11, pp. 2825–2832, Nov. 2004.

[112] G. Orecchini *et al.*, "Green technologies and RFID: Present and future," *Appl. Comput. Electromagn. Soc. J.*, vol. 25, no. 3, pp. 230–238, Mar. 2010.

[113] S. Steudel *et al.*, "Comparison of organic diode structures regarding high-frequency rectification behavior in radio-frequency identification tags," *J. Appl. Phy.*, vol. 99, no. 11, Jun. 2006, 114519.

[114] L. Valentini and J. Kenny, "Novel approaches to developing carbon nanotube based polymer composites: Fundamental studies and nanotech applications," *Polymer*, vol. 46, no. 17, pp. 6715–6718, Aug. 2005.

[115] V. Marinov *et al.*, "Direct-write vapor sensors on FR4 plastic substrates," *IEEE Sensor J.*, vol. 7, no. 6, pp. 937–944, Jun. 2007.

[116] T. Unander and H.-E. Nilsson, "Characterization of printed moisture sensors in packaging surveillance applications," *IEEE Sensor J.*, vol. 9, no. 8, pp. 922–928, Aug. 2009.

[117] S. Couderc, B. Kim, and T. Someya, "Cellulose-based composite as a raw material for flexible ANS ultra-lightweight mechanical switch devices," in *Proc. IEEE 22nd Int. Conf. Micro Electro Mech. Syst.*, Sorrento, Italy, Jan. 2009, pp. 646–649.

[118] P. F. Goldsmith, "Quasioptics in millimeterwave systems," in *Proc. 12th Eur. Microw. Conf.*, Sep. 1982, pp. 16–24.

[119] P. Goldsmith, "Quasi-optical techniques," *Proc. IEEE*, vol. 80, no. 11, pp. 1729–1747, Nov. 1992.

[120] P. F. Goldsmith, *Quasioptical Systems*. New York, NY, USA: IEEE Press, 1998.

[121] T. Djerafi, K. Wu, and S. Tatu, "3 dB 90° hybrid quasi-optical coupler with air field slab in SIW technology," *IEEE Microw. Wireless Compon. Lett.*, vol. 24, no. 4, pp. 221–223, Apr. 2014.

[122] A. Gonzalez, Y. Fujii, K. Kaneko, and Y. Uzawa, "Quasi-optical attenuators for alma band 10 receiver," in *Proc. Asia-Pacific Microw. Conf. Proc.*, Dec. 2011, pp. 1977–1980.

[123] K. Song, F. Zhang, S. Hu, and Y. Fan, "Millimeter-wave-wave quasi-optical low-loss power combiner based on dipole antenna," *Electron. Lett.*, vol. 49, no. 18, pp. 1160–1162, Aug. 2013.

[124] S. Kawasaki and T. Itoh, "Quasi-optical planar arrays with fets and slots," *IEEE Trans. Microw. Theory Tech.*, vol. 41, no. 10, pp. 1838–1844, Oct. 1993.

[125] S. Hollung, W. Shiroma, M. Markovic, and Z. Popovic, "A quasi-optical isolator," *IEEE Microw. Guided Wave Lett.*, vol. 6, no. 5, pp. 205–206, May 1996.

[126] N. Gagnon and J. Shaker, "Accurate phase measurement of passive non-reciprocal quasi-optical components," *IEE Proc.—Microw. Antennas Propag.*, vol. 152, no. 2, pp. 111–116, Apr. 2005.

[127] R. Dickie, R. Cahill, V. Fusco, H. Gamble, and N. Mitchell, "THz frequency selective surface filters for earth observation remote sensing instruments," *IEEE Trans. Terahertz Sci. Technol.*, vol. 1, no. 2, pp. 450–461, Nov. 2011.

[128] M. Euler and V. Fusco, "Quasi optical circulator X-band scale model utilizing a circular polarization frequency selective surface," in *Proc. 4th Eur. Conf. Antennas Propag.*, Apr. 2010, pp. 1–4.

[129] W. Shiroma, S. Bundy, S. Hollung, B. Bauernfeind, and Z. Popovic, "Cascaded active and passive quasi-optical grids," *IEEE Trans. Microw. Theory Tech.*, vol. 43, no. 12, pp. 2904–2909, Dec. 1995.

[130] J. Birkeland and T. Itoh, "A 16 element quasi-optical fet oscillator power combining array with external injection locking," *IEEE Trans. Microw. Theory Tech.*, vol. 40, no. 3, pp. 475–481, Mar. 1992.

[131] M. Forman, T. Marshall, and Z. Popovic, "Two Ka-band quasi-optical amplifier arrays," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 1999, vol. 4, pp. 1829–1832.

[132] S. Hollung, A. Cox, and Z. Popovic, "A bi-directional quasi-optical lens amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. 45, no. 12, pp. 2352–2357, Dec. 1997.

[133] J. Hacker, R. Weikle, M. Kim, P. De Lisio, and D. Rutledge, "A 100-element planar Schottky diode grid mixer," *IEEE Trans. Microw. Theory Tech.*, vol. 40, no. 3, pp. 557–562, Mar. 1992.

[134] G. Chattopadhyay, D. Miller, H. LeDuc, and J. Zmuidzinas, "A dual-polarized quasi-optical SIS mixer at 550 GHz," *IEEE Trans. Microw. Theory Tech.*, vol. 48, no. 10, pp. 1680–1686, Oct. 2000.

[135] V. Bezborodov, V. Kiseliov, Y. Kuleshov, and M. Yanovsky, "A terahertz quasi-optical polarization phase shifter with the crystal quartz phase sections," in *Proc. 5th Int. Kharkov Symp. Phys. Eng. Microw. Millimeter Submillimeter Waves*, Jun. 2004, vol. 2, pp. 603–605.

[136] G. Denisov *et al.*, "Fast quasi-optical phase shifter based on induced photoconductivity in silicon," in *Proc. Joint 32nd Int. Conf. Infrared Millimeter Waves/15th Int. Conf. Terahertz Electron.*, Sep. 2007, pp. 795–796.

[137] N. Camilleri, "A quasi-optical multiplying slot array," *IEEE Trans. Microw. Theory Tech.*, vol. 33, no. 11, pp. 1189–1195, Nov. 1985.

[138] A. Moussessian *et al.*, "A terahertz grid frequency doubler," *IEEE Trans. Microw. Theory Tech.*, vol. 46, no. 11, pp. 1976–1981, Nov. 1998.

[139] M. Cryan *et al.*, "Simulation and measurement of quasi-optical multipliers," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 3, pp. 451–464, Mar. 2001.

[140] F. Jiang *et al.*, "High-speed monolithic millimeter-wave switch array," *IEEE Microw. Guided Wave Lett.*, vol. 8, no. 3, pp. 112–114, Mar. 1998.

[141] F. Alimenti *et al.*, "Crossed dipole frequency doubling RFID TAG based on paper substrate and ink-jet printing technology," in *Proc. IEEE MTT-S Int. Microw. Symp. Dig.*, May 2010, pp. 840–842.

[142] M. Virili *et al.*, "7.5–15 MHz organic frequency doubler made with pentacene-based diode and paper substrate," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Tampa, FL, USA, Jun. 2014, DOI: 10.1109/MWSYM.2014.6848395.

[143] L. Roselli *et al.*, "Feasibility study of a fully organic frequency doubler for harmonic RFID applications," in *Proc. IEEE 12th Top. Meeting Silicon Monolithic Integr. Circuits RF Syst.*, Jan. 2012, pp. 203–206.

[144] R. Gonalves *et al.*, "Smart floor: Indoor navigation based on RFID," in *Proc.*

IEEE Wireless Power Transfer, May 2013, pp. 103–106.

[145] G. Orecchini, L. Yang, M. Tentzeris, and L. Roselli, "Wearable battery-free active paper printed RFID tag with human-energy scavenger," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2011, DOI: 10.1109/MWSYM.2011.5973290.

[146] C.-R. Yu, C.-L. Wu, C.-H. Lu, and L.-C. Fu, "Human localization via multi-cameras and floor sensors in smart home," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Oct. 2006, vol. 5, pp. 3822–3827.

[147] S. Chang, S. Ham, S. Kim, D. Suh, and H. Kim, "Ubi-floor: Design and pilot implementation of an interactive floor system," in *Proc. 2nd Int. Conf. Intell. Human-Mach. Syst. Cybern.*, Aug. 2010, vol. 2, pp. 290–293.

[148] J. Morgado and A. Konig, "Low-power concept and prototype of distributed resistive pressure sensor array for smart floor and surfaces in intelligent environments," in *Proc. 9th Int. Multi-Conf. Syst. Signals Devices*, Mar. 2012, DOI: 10.1109/SSD.2012.6198020.

[149] J. Zhao and Y. Wang, "Autonomous ultrasonic indoor tracking system," in *Proc. Int. Symp. Parallel Distrib. Process. Appl.*, Dec. 2008, pp. 532–539.

[150] R. Imura, "The role of networked RFID for driving the ubiquitous," in *Proc. IEEE Conf. Emerging Technol. Factory Autom.*, Sep. 2006, pp. 1115–1118.

[151] Y. Tan, K. Y. Hoe, and S. Panda, "Energy harvesting using piezoelectric igniter for self-powered radio frequency (RF) wireless sensors," in *Proc. IEEE Int. Conf. Ind. Technol.*, Dec. 2006, pp. 1711–1716.

[152] Linx Technologies, "KH2 series transmitter/encoder data guide." [Online]. Available: https://www.linxtechnologies.com/resources/data-guides/txe-xxx-kh2.pdf

[153] D. Smith, D. Miniutti, T. Lamahewa, and L. Hanlen, "Propagation models for body-area networks: A survey and new outlook," *IEEE Antennas Propag. Mag.*, vol. 55, no. 5, pp. 97–117, Oct. 2013.

[154] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.6: Wireless Body Area Networks*, IEEE Std. 802.15.6-2012, Feb. 2012.

[155] J. Thomas and M. Qidwai, "The design and application of multifunctional structure-battery materials systems," *JOM*, vol. 57, no. 3, pp. 18–24, Mar. 2005.

[156] A. Bonfiglio and D. De Rossi, *Wearable Monitoring Systems*. New York, NY, USA: Springer-Verlag, 2011.

[157] S. Lee and B. Youn, "A design and experimental verification methodology for an energy harvester skin structure," *Smart Mater. Struct.*, vol. 20, no. 5, May 2011, DOI: 10.1088/0964-1726/20/5/057001.

[158] R. Vullers, R. van Schaijk, I. Doms, C. van Hoof, and R. Mertens, "Microwave energy harvesting," *Solid-State Electron.*, vol. 53, no. 4, pp. 684–693, Apr. 2009.

[159] P. Jaffe and J. McSpadden, "Energy conversion and transmission modules for space solar power," *Proc. IEEE*, vol. 101, no. 6, pp. 1424–1437, Jun. 2013.

[160] J. McSpadden and J. Mankins, "Space solar power programs and microwave wireless power transmission technology," *IEEE Microw. Mag.*, vol. 3, no. 4, pp. 46–57, Dec. 2002.

ABOUT THE AUTHORS

Luca Roselli (Senior Member, IEEE) was born in Florence, Italy, in 1962. He received the "Laurea" degree in electronic engineering from the University of Florence, Florence, Italy, in 1988.

From 1988 to 1991, he worked at the University of Florence on SAW devices. In November 1991, he joined the University of Perugia, Perugia, Italy, where he is currently working as an Associate Professor and where he has been teaching several classes on electronic devices, microwave electronics, high-frequency (HF) electronic components, and applied electronics. Since 2000, he has been coordinating research activity at the High Frequency Electronics (HFE) Laboratory. In that year, he also founded the spin-off company WiS (Wireless Solutions) Srl, operating in the field of microwave electronic systems with which he cooperated as a consultant until it joined the new company ART Srl. in 2008. From 2008 to 2012, he was the Director of the Technical and Scientific Committee of ART Srl and a member of the Board of Directors. In 2005, he founded a second spin-off company: DiES (Digital Electronic Solutions) Srl. His research interests mainly focus on the design of high-frequency electronic circuits and systems, including the development of numerical methods for electronic circuit analysis with special attention to RFID–NFC systems, new materials (including organic and recyclable ones), and far-field wireless power transfer. In these fields, he published more than 220 contributions in international reviews and peer-reviewed conferences, the interest in which is testified by an HF index of 20 (source Google Scholar) and more than 1350 citations.

Mr. Roselli was the Chairman of the VII Computational Electromagnetic in Time Domain Workshop in 2007. In 2013, he was the Chairman of the First IEEE Wireless Power Transfer Conference (WPTC). Currently, he is member of the list of experts of Italian Ministry of Research and University (MIUR); member of several IEEE Technical Committees [MTT-24 RFID Technologies (past chair), MTT-25 RF nanotechnolgies, MTT-26 Wireless Power Transfer]; member of the Sub Committee 32 RFID Technologies of International Microwave Symposium (IMS) (past chair); member of the European Research Council (ERC) Panel PE7; member of the Advisory Committee of IEEE WPTC. He is involved in the boards of several international conference (RWCOM, RFID-TA, EuCAP, MAREW). He is a reviewer for many international conferences and journals (the Proceedings of the IEEE, the IEEE Transactions on Microwave Theory and Techniques, IEEE Microwave and Wireless Components Letters, *ACES Journal*, *Radioengineering Journal*, Hindawi publishing corporation, *Elsevier Organic Electronics*, *ASP Nanoscience and Nanotechnology Letters*).

Nuno Borges Carvalho (Senior Member, IEEE) was born in Luanda, Angola, in 1972. He received the Diploma and Doctoral degrees in electronics and telecommunications engineering from the University of Aveiro, Aveiro, Portugal, in 1995 and 2000, respectively.

He is a Full Professor at the Universidade de Aveiro, Aveiro, Portugal and a Senior Research Scientist with the Instituto de Telecomunicações (IT), Universidade de Aveiro, where he coordinates the wireless communication thematic area and the Radio Systems Group. He coauthored *Intermodulation in Microwave and Wireless Circuits* (Reading, MA, USA: Artech House, 2003) and *Microwave and Wireless Measurement Techniques* (Cambridge, U.K.: Cambridge Univ. Press, 2013). He has been a reviewer and author of over 200 papers in magazines and conferences. He coholds four patents. His main research interests include software-defined radio front–ends, wireless power transmission, nonlinear distortion analysis in microwave/wireless circuits and systems, and measurement of nonlinear phenomena. He has recently been involved in the design of dedicated radios and systems for newly emerging wireless technologies.

Dr. Borges Carvalho is the Chair of the IEEE MTT-11 Technical Committee. He is a member of IEEE MTT-20, MTT-24, and MTT-26 and the IEEE Portugal Section Chair. He is an Associate Editor for the IEEE Transactions on Microwave Theory and Techniques, IEEE Microwave Magazine, *Cambridge Journal on Wireless Power Transmission* and the Chair of the URSI–Portugal Metrology Group.

Federico Alimenti (Senior Member, IEEE) was born in Foligno, Italy, in 1968. He received the Laurea degree (*magna cum laude*) and the Ph.D. degree from the University of Perugia, Perugia, Italy, in 1993 and 1997 respectively, both in electronic engineering.
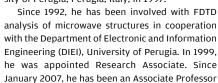
In 1993, he held an internship at Daimler-Benz Aerospace, Ulm, Germany. In 1996, he was recipient of the URSI young scientist award and visiting scientist at the Technical University of Munich, Munich, Germany. Since 2001, he has been with the Department of Engineering, University of Perugia, as an Aggregate Professor, teaching the class of microwave electronics. Between 2006 and 2010, he was a member of the Academic Senate in the same university. He authored more than 150 papers in refereed journals and conference proceedings. He coauthored several book chapters and a European patent (EP13161946.2) on zero-power wireless sensors. His scientific interests concern the design and the experimental characterization of radio-frequency (RF)-integrated circuits in complementary metal–oxide–semiconductor (CMOS) and BiCMOS technologies.

Prof. Alimenti was a recipient of the IET Premium (Best Paper) Award in 2013 and was the Technical Program Committee (TPC) Chair of the IEEE Wireless Power Transfer (WPT) Conference. He has been a local coordinator of the ARTEMOS project, ENIAC, call 3, 2010. His H-index is equal to 14 (source: Google Scholar).

Paolo Mezzanotte (Member, IEEE) was born in Perugia, Italy, in 1965. He received the Ph.D. degree in electrical engineering from the University of Perugia, Perugia, Italy, in 1997.

Since 1992, he has been involved with FDTD analysis of microwave structures in cooperation with the Department of Electronic and Information Engineering (DIEI), University of Perugia. In 1999, he was appointed Research Associate. Since January 2007, he has been an Associate Professor with the same university, teaching the classes on radio-frequency engineering. His research activities concern numerical methods and computer-aided design (CAD) techniques for passive microwave structures and the analysis and design of microwave and millimeter-wave circuits. More recently his research interests have mainly focused on the study of advanced technologies such as LTCC, RF-MEMS, and microwave circuits printed on green substrates. These research activities are testified by more than 100 publications in the most important specialized journals and at the main conferences of the microwave scientific community.

Prof. Mezzanotte serves as a Reviewer of a number of IEEE journals and belongs to the Top Amplifier Research Group in a European Team (TARGET) and to the Advanced MEMS for RF and Millimeterwave Communications (AMICOM) Networks of Excellence. He is a member of the IMS Technical Program Review Committee and of the Technical Committee MTT-24 RFID Technologies. He serves as an Associate Editor of the *ACES Journal*. He was the Co-Chairman of the Seventh International Workshop on Computational Electromagnetics in Time Domain (CEM–TD 2007). His present H-index (ISI journals) is equal to 13.

Giulia Orecchini received the Laurea degree in electronic engineering and the Ph.D. degree in electronic and information engineering from the University of Perugia, Perugia, Italy, in 2008 and 2012, respectively.
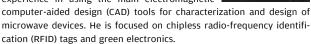
During her Ph.D. studies she joined the ATHENA Research Group, Georgia Institute of Technology, Atlanta, GA, USA. Her research interests concern the development of radio-frequency identification (RFID) electronic systems and technologies. She is currently working as a Research Assistant at the University of Perugia.

Marco Virili (Student Member, IEEE) was born in Terni, Italy, in 1983. He received the Laurea degree in electronic engineering from the University of Perugia, Perugia, Italy, in 2009, where he is currently working toward the Ph.D. degree at the Department of Engineering.

His scientific interests concern the design and the experimental validation of radio-frequency (RF) and microwave circuits. He has a strong experience in using the main electromagnetic computer-aided design (CAD) tools for characterization and design of microwave devices. He is focused on chipless radio-frequency identification (RFID) tags and green electronics.

Chiara Mariotti (Student Member, IEEE) was born in Assisi, Italy, in 1987. She received the Laurea degree (*magna cum laude*) in electronic and telecommunication engineering from the University of Perugia, Perugia, Italy, in 2011, where she is currently working toward the Ph.D. degree at the High Frequency Electronics (HFE) Laboratory, working on green technologies for passive and energetically autonomous radio-frequency identification (RFID) tags.

In Spring 2012, she joined the ATHENA Research Group, Georgia Institute of Technology, Atlanta, GA, USA, for six months, working on eco-compatible indoor localization systems and other types of sensors fabricated by means of inkjet printing technology. At the end of 2013, she again joined the ATHENA Research Group for six months to work on multilayer inkjet printed systems and devices such as passives and microfluidics sensors.

Ricardo Gonçalves (Student Member, IEEE) was born in Lisbon, Portugal, in 1988. He received the B.Sc. and M.Sc. (*magna cum laude*) degrees in electronics and telecommunications engineering from the Instituto Superior de Engenharia de Lisboa, Lisbon, Portugal, in 2010 and 2012, respectively. He is currently working toward the Ph.D. degree in electrical engineering at the University of Aveiro, Aveiro, Portugal.

He is a Researcher at the Instituto de Telecomunicações, Aveiro, Portugal. In 2011, he was awarded with merit scholarship for exceptional academic performances. From 2010 to 2012, he worked at Dailywork I&D as a Developer Engineer in Embedded Electronic Systems for Vehicular Location Systems and Healthcare Platforms. His main research interests include wireless power transfer systems, radio-frequency identification (RFID), and wireless passive sensor networks. His current focus is on the development of printed antennas and microwave circuits in nonconventional materials such as textiles, ceramics, plastic, paper, and cork, for the former applications.

Pedro Pinho (Member, IEEE) was born in Vale de Cambra, Portugal, in 1974. He received the Licenciado and M.S. degrees in electrical and telecommunications engineering and the Ph.D. degree in electrical engineering from the University of Aveiro, Aveiro, Portugal, in 1995 and 2000, respectively. from the University of Aveiro, Aveiro, Portugal, in 1997, 2000, and 2004, respectively.

He is currently a Professor Adjunto at the Department of Electrical Telecommunications and Computers Engineering, Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisbon, Portugal, and a Member of the Research Staff at the Institute for Telecommunications, Aveiro, Portugal (since 1997). He authored or coauthored more than 90 papers in conferences and international journals and four book chapters. His current research interest is in antennas for location systems, reconfigurable antennas, and antenna design for passive sensors in nonconventional materials.
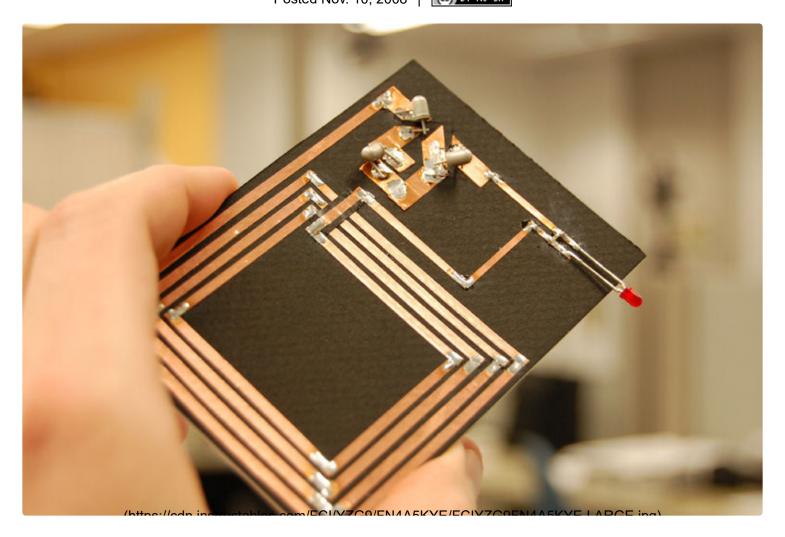
RFID READER DETECTOR AND TILT-SENSITIVE RFID TAG
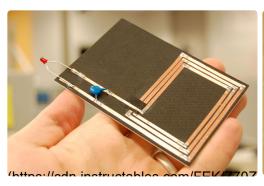
Technology (/technology/) > Science | by nmarquardt (/member/nmarquardt/)   Follow

286,565   💜 531   157
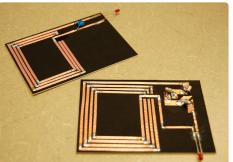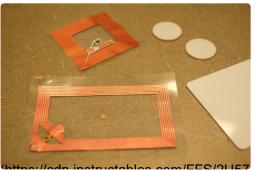
Posted Nov. 10, 2008 | (cc) BY-NC-SA

(https://cdn.instructables.com/FCIYZG9/FN4A5KYF/FCIYZG9FN4A5KYF.LARGE.jpg)

(https://cdn.instructables.com/FFK7Z0Z (https://cdn.instructables.com/FW3/I (https://cdn.instructables.com/FFS/3H5Z

[The 'rub' (http://www.usingenglish.com/reference/idioms/there%27s+the+rub.html)](http://www.usingenglish.com/reference/idioms/there%27s+the+rub.html)
Want to detect the presence of RFID readers? Want to control when a RFID tag is active or readable? We describe how to do both using bits of copper and card, and some readily available electronics hardware.

Longer preamble
Radio frequency identification ([RFID (http://en.wikipedia.org/wiki/RFID)](http://en.wikipedia.org/wiki/RFID)) is rapidly growing in popularity. RFID *tags* are found everywhere. They're attached to container freight, in those funny-looking white labels you find in newly purchased books, embedded in many corporate ID cards and passports, etc. The tags have a few common properties: they transmit a unique ID number, are optimized to be 'read' from predefined distances, and are usually small so they can remain unobtrusive or hidden.

RFID *readers* are used to track nearby tags by wirelessly reading a tag's unique ID (see Figure 4); a tag simply has to be brought into physical proximity with a reader to be read. Readers are mostly used for industrial or commercial purposes, e.g. asset tracking or electronic payment. Wal-mart use RFID tags and readers in their supply chain. The technology is also used in mass transit systems in cities like [London (http://en.wikipedia.org/wiki/Oyster_card)](http://en.wikipedia.org/wiki/Oyster_card) and [Hong Kong (http://en.wikipedia.org/wiki/Octopus_card)](http://en.wikipedia.org/wiki/Octopus_card). In Japan, many mobile phones incorporate readers to enable e-money payments in shops and vending machines.
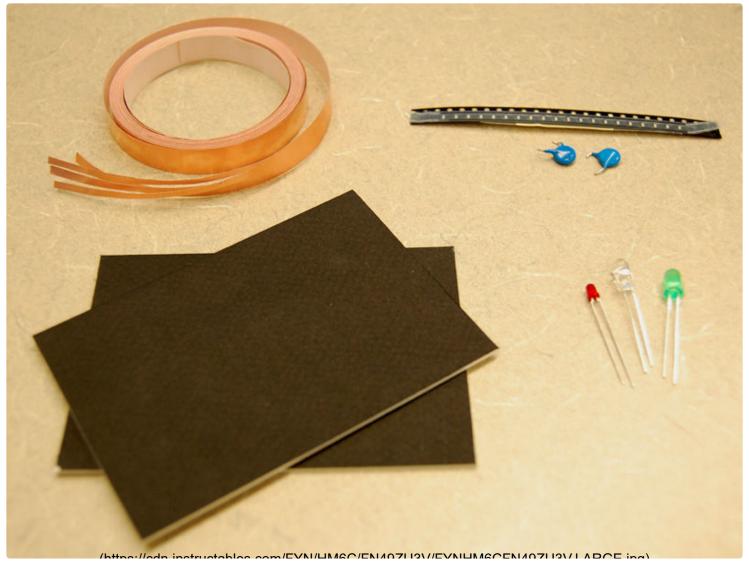
For those of us who want to experiment with RFID, the problem is that the technology is almost always ☐[black boxed (http://en.wikipedia.org/wiki/Black_box)](http://en.wikipedia.org/wiki/Black_box).

That is, the inner workings of a tag and its interaction with a reader is hidden from view, and thus difficult to have much control over.

In the two exercises that follow (building a RFID reader detector (https://www.instructables.com/id/SX28QZ1FN4H8QCG/) and a tilt-sensitive RFID tag (https://www.instructables.com/id/SGW3J0SFN49WUIF/)), we offer an example of how you can start revealing some of the workings of RFID and thus gain some control over the technology. The two exercises also hopefully show that the technology is relatively simple and how it can be extended to support some interesting interactions. We offer some other possibilities that build on our examples at the end (https://www.instructables.com/id/SQKIRD4FN49ZUS1/).

Add Tip   Ask Question
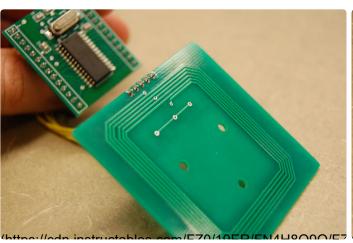
Step 1: Material and Tools
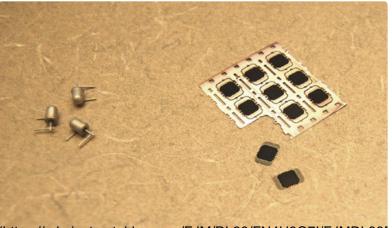
This section provides an overview of the necessary materials and tools.

Materials (see Figure 1):

We need the following material to built the basic RFID reader detector.

- Cardboard (around 100x70 mm)

- Conductive copper tape (e.g., order number 1218478 at www.farnell.com)

- Capacitor 82 pF (picofarad) (e.g., order number 1138852 at www.farnell.com)

- Low current LED (light-emitting diode) (e.g., order number 1003207at www.farnell.com)

Tools (see Figure 2 and 3):

- Craft knife and scissors

- Insulating tape (e.g., order number 1373979 at www.farnell.com)

- Soldering iron and solder

RFID reader for testing (see Figure 4):

To test our RFID tags we need an RFID reader that can operate at a frequency of 13.56 MHz.

There many readers for this widely used RFID standard, for instance the Sonmicro MIFARE USB reader (http://www.sonmicro.com/).

Note: The Phidget RFID reader does *not* work with the tags created in this project, as it uses a different frequency for communication with the tags (125 kHz).

Advanced material (see Figure 5):
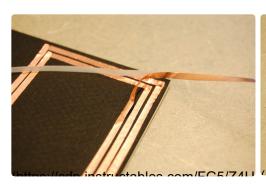
The following material is necessary to build the second part of the project: the tilt-sensitive RFID tag.
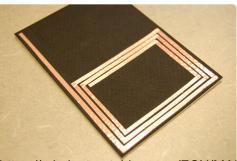
- Micro tilt switches (e.g., www.digikey.com)

- RFID ICs (e.g., MIFARE Standard 1k, part no. 568-2219-1-ND at www.digikey.com)
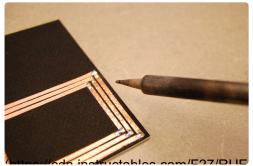
Add Tip Ask Question

Step 2: Building the RFID Antenna

(https://cdn.instructables.com/FGI0OXD/FN4A5N20/FGI0OXDFN4A5N20.LARGE.jpg)

This step describes how to build the antenna for the RFID tag.

Building the RFID tag antenna
To build the tag's antenna follow these three steps.
1. Cut the conductive copper tape into thin stripes of around 2mm (see Figure 1).
2. Tape these stripes (see Figure 2) in loops around one half of the cardboard (see Figure 3 for the layout of the antenna). The tag should have between 3-4 loops for the antenna.
3. Solder all the connections between the copper tape. Sometimes, this isn't necessary as the tape's adhesive backing is conductive, but solder the connections if you want to be on the safe side.

Now we have created our RFID tag antenna, and we will add the "RFID reader detection" functionality in the following step.
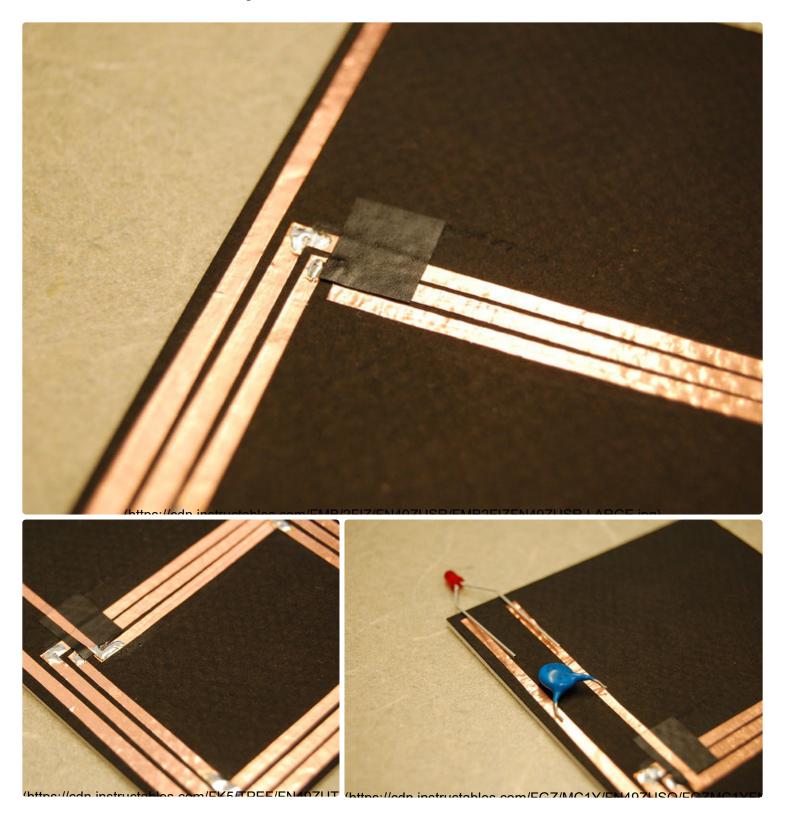
A little background
RFID readers transmit an electromagnetic (EM) field with their reader antenna. This EM field induces a current in the antenna for all RFID tags within reading distance. This induced current activates the RFID chip that is connected to the tag's antenna. This chip then modulates a response (usually the unique ID number) that is transmitted back to the reader. The antenna of an RFID tag is usually a thin copper wire that is arranged in loops. The loops allow the emitted EM field of the RFID reader to induce current to the antenna of the tag.
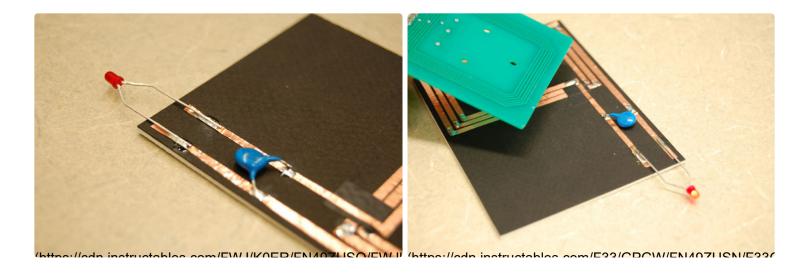
Add Tip | Ask Question

Step 3: RFID Reader Detection

This step describes how to add a simple mechanism to the RFID tag antenna that allows us detect nearby RFID readers.

Antenna connection

First, we add a small piece of insulation tape for the connection of the inner end of the antenna loop (as illustrated in Figure 1). This is to insulate the outer loops. Then we add another copper tape strip to the inner end of the antenna as shown in Figure 2. Here again we solder the two ends of the conductive copper tape together.

Capacitor and LED

Next, we add the capacitor (82 pF) and the low current LED to the tag as shown in Figure 3. They are connected in parallel. We also solder these two components to the copper tape (see Figure 4).
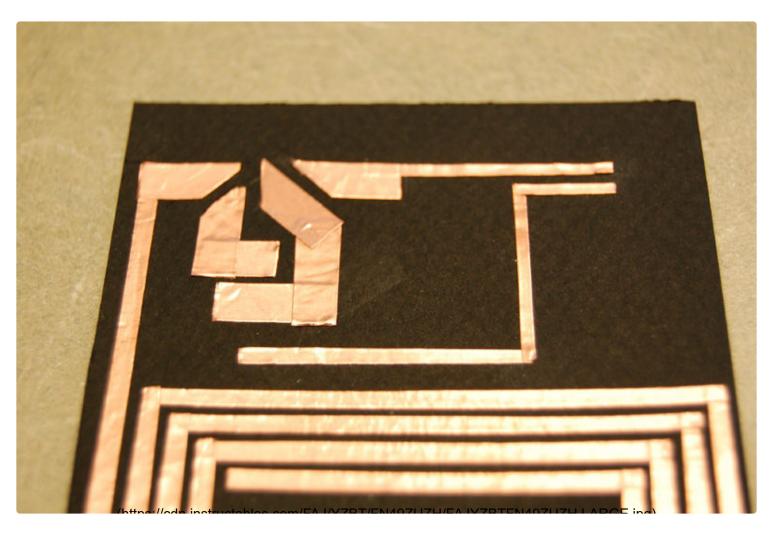
Testing
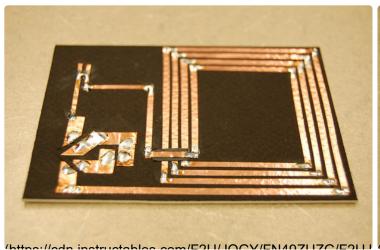
With these simple steps, our RFID reader detector is finished! By bringing our DIY RFID detector close to an RFID reader (as shown in Figure 5), the connected LED lights up. With the Sonmicro reader hardware the distance to the reader has to be below 8-10 cm; however, there are RFID readers available with a stronger EM field and therefore a higher maximum reading distance.
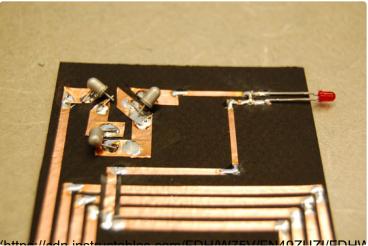
In the next step of the instructable we will show how to extend a basic RFID tag and make it tilt-sensitive.

Add Tip Ask Question

Step 4: Tilt-Sensitive RFID Tag
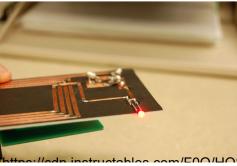
(https://cdn.instructables.com/FAJYZRT/FN49ZHZH/FAJYZRTFN49ZHZH.LARGE.jpg)
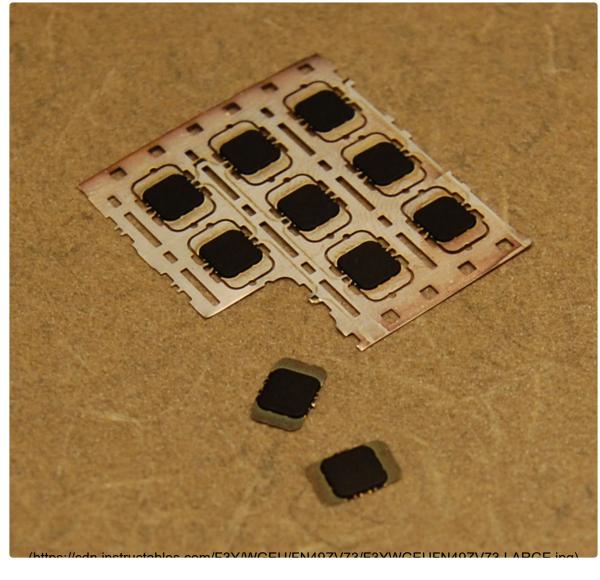
We now describe the process of how to build a tilt-sensitive RFID tag. This extends the previous exercise.

Antenna

The antenna for this second RFID tag is similar to the first antenna we built. We thus need another piece of cardboard and to repeat the steps described earlier in STEP 2 (https://www.instructables.com/id/S2YO8NWFN4H8Q8A/) of this instructable.

Tilt-sensitive tag

Next, we add additional copper tape connections to the tag, as shown in Figure 1. These connections allow us to connect three tilt switches, a capacitor, and the LED to the antenna. Again, all the connections of the copper tape are soldered together. We add the three tilt switches to the tag as shown in Figure 3. The tilt switches are soldered to the copper tape, and it is important to connect them in a slight angle (around 5-10 degrees) as shown in Figure 4. This makes sure that the silt switches are in a *closed* state while the RFID tag is in a horizontal position, and in a *open* state while the tag is in a vertical position.

Again, we also add an LED and a capacitor to the antenna as shown in Figure 3 (we use a different form factor of the capacitor here just to illustrate the alternative options).

Testing the tilt-sensitive tag

We can now use our Sonmicro RFID reader again to test our new tilt-sensitive RFID tag. The tag is activate while in a horizontal position as in Figure 5, and is inactive when in a vertical position as in Figure 6.

Using RFID chips

We can now replace the connected capacitor and LED from our tag with an RFID

MICROWAVE AND RADIO FREQUENCY ENGINEERING

INDEX

TRANSMISSION LINES

TELEGRAPHER'S EQUATIONS

$$(1) \quad \frac{\partial V}{\partial z} = -L \frac{\partial I}{\partial t} \qquad (2) \quad \frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t}$$

By taking the partial derivative with respect to *z* of equation 1 and partial with respect to *t* of equation 2, we can get:

$$(i) \quad \frac{\partial^2 V}{\partial z^2} = LC \frac{\partial^2 V}{\partial t^2} \qquad (ii) \quad \frac{\partial^2 I}{\partial z^2} = LC \frac{\partial^2 I}{\partial t^2}$$

SOLVING THE EQUATIONS

To solve the equations (i) and (ii) above, we guess that $F(u) = F(z \pm vt)$ is a solution to the equations. It is found that the unknown constant *v* is the wave propagation velocity.

$$\boxed{V_{total} = V_+ (z - vt) + V_- (v + vt)} \quad \text{where:}$$

z is the position along the transmission line, where the load is at *z*=0 and the source is at *z*=-*l*, with *l* the length of the line.

v is the **velocity of propagation** $1/\sqrt{LC}$ or ω/β, the speed at which the waveform moves down the line; see p 2

t is time

THE COMPLEX WAVE EQUATION

The general solutions of equations (i) and (ii) above yield the complex wave equations for voltage and current. These are applicable when the excitation is sinusoidal and the circuit is under steady state conditions.

$$\boxed{V(z) = V^+ e^{-j\beta z} + V^- e^{+j\beta z}}$$

$$\boxed{I(z) = I^+ e^{-j\beta z} + I^- e^{+j\beta z}}$$

$$\boxed{I(z) = \frac{V^+ e^{-j\beta z} + V^- e^{+j\beta z}}{Z_0}} \quad \text{where:}$$
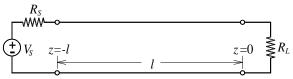
$e^{-j\beta z}$ and $e^{+j\beta z}$ represent wave propagation in the +*z* and −*z* directions respectively,

$\beta = \omega\sqrt{LC} = \omega/v$ is the **phase constant,**

$Z_0 = \sqrt{L/C}$ is the **characteristic impedance** of the line.

These equations represent the voltage and current **phasors**.

+/− WATCHING SIGNS

By convention *z* is the variable used to describe position along a transmission line with the origin *z*=0 set at the load so that all other points along the line are described by **negative** position values.

Ohm's law for right- and left-traveling disturbances:
$$V_+ = I_+ Z_0 \qquad\qquad V_- = -I_- Z_0$$

v_p VELOCITY OF PROPAGATION [cm/s]

The velocity of propagation is the speed at which a wave moves down a transmission line. The velocity approaches the speed of light but may not exceed the speed of light since this is the maximum speed at which information can be transmitted. But v_p may exceed the speed of light mathematically in some calculations.

$$\boxed{v_p = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{\omega}{\beta}} \quad \text{where:}$$

L = inductance per unit length [H/cm]
C = capacitance per unit length [F/cm]
ε = permittivity of the material [F/cm]
μ = permeability of the material [H/cm]
ω = frequency [radians/second]
β = phase constant

Phase Velocity The velocity of propagation of a TEM wave may also be referred to as the *phase velocity*. The phase velocity of a TEM wave in conducting material may be described by:

$$\boxed{v_p = \omega\delta = \frac{\omega}{k} = c \frac{2\pi\delta}{\lambda_0} = c \frac{1}{\sqrt{\varepsilon_{r\ \text{eff}}}}} \quad \text{where:}$$

δ = skin depth [m]
c = speed of light 2.998×10^8 m/s
λ_0 = wavelength in the material [m]

Z_0 CHARACTERISTIC IMPEDANCE   [Ω]

The **characteristic impedance** is the resistance initially *seen* when a signal is applied to the line. It is a physical characteristic resulting from the materials and geometry of the line.

Lossless line: $\quad Z_0 \equiv \sqrt{\dfrac{L}{C}} = \dfrac{V_+}{I_+} = -\dfrac{V_-}{I_-}$

Lossy line: $\quad Z_0 \equiv \sqrt{\dfrac{R + j\omega L}{G + j\omega C}} = |Z_0|e^{j\phi_z}$

L = inductance per unit length [H/cm]
C = capacitance per unit length [F/cm]
V_+ = the forward-traveling (left to right) voltage [V]
I_+ = the forward-traveling (left to right) current [I]
V_- = the reverse-traveling (right to left) voltage [V]
I_- = the reverse-traveling (right to left) current [I]
R = the line resistance per unit length [Ω/cm]
G = the conductance per unit length [Ω^{-1}/cm]
ϕ = phase angle of the complex impedance [radians]

y_0 CHARACTERISTIC ADMITTANCE   [Ω^{-1}]

The **characteristic admittance** is the reciprocal of the characteristic impedance.

$$y_0 \equiv \sqrt{\dfrac{C}{L}} = \dfrac{I_+}{V_+} = -\dfrac{I_-}{V_-}$$

r REFLECTION COEFFICIENT

The **reflection coefficient** is the ratio of reflected voltage to the forward-traveling voltage, a value ranging from –1 to +1 which, when multiplied by the wave voltage, determines the amount of voltage reflected at one end of the transmission line.

$$\rho \equiv \dfrac{V_-}{V_+} = -\dfrac{I_-}{I_+}$$

A reflection coefficient is present at each end of the transmission line:

$$\rho_{\text{source}} = \dfrac{R_S - z_0}{R_S + z_0} \qquad \rho_{\text{load}} = \dfrac{R_L - z_0}{R_L + z_0}$$

t TRANSMISSION COEFFICIENT

The **transmission coefficient** is the ratio of total voltage to the forward-traveling voltage, a value ranging from 0 to 2.

$$\tau \equiv \dfrac{V_{total}}{V_+} = 1 + \rho$$

TOF TIME OF FLIGHT   [s]

The **time of flight** is how long it takes a signal to travel the length of the transmission line

$$TOF \equiv \dfrac{l}{v} = l\sqrt{LC} = \sqrt{L_{TOT}C_{TOT}}$$

l = length of the transmission line [cm]
v = the velocity of propagation $1/\sqrt{LC}$, the speed at which the waveform moves down the line
L = inductance per unit length [H/cm]
C = capacitance per unit length [F/cm]
L_{TOT} = total inductance [H]
C_{TOT} = total capacitance [F]

DERIVED EQUATIONS

$$V_+ = z_0 I_+ = \left(V_{TOT} + I_{TOT}z_0\right)/2$$
$$V_- = -z_0 I_- = \left(V_{TOT} - I_{TOT}z_0\right)/2$$
$$I_+ = y_0 V_+ = \left(I_{TOT} + V_{TOT}y_0\right)/2$$
$$I_- = -y_0 V_- = \left(I_{TOT} - V_{TOT}y_0\right)/2$$

C_n FOURIER SERIES

The function $x(t)$ must be periodic in order to employ the Fourier series. The following is the exponential Fourier series, which involves simpler calculations than other forms but is not as easy to visualize as the trigonometric forms.

$$C_n = \dfrac{1}{T}\int_{t_1}^{t_1 + T} x(t)e^{-jn\omega_0 t}\, dt$$

C_n = amplitude
T = period [s]
t = time [s]
n = the harmonic (an integer)
w_0 = frequency $2\pi/T$ [radians]

The function $x(t)$ may be delayed in time. All this does in a Fourier series is to shift the phase. If you know the C_ns for $x(t)$, then the C_ns for $x(t\text{-}\alpha)$ are just $C_n e^{-jn\omega 0\alpha}$. (Here, C_ns is just the plural of C_n.)

<div style="border:1px solid">

C CAPACITANCE [F]

$$v(t) = \frac{1}{C}\int_0^t i\,d\tau + v(0) \qquad I_{cap} = C\frac{dV_{cap}}{dt}$$

$$\boxed{v(t) = v_f + \left(v_0 - v_f\right)e^{-t/t}} \qquad i(t) = i_f + \left(i_0 - i_f\right)e^{-t/t}$$

$$P(t) = i_0^{\;2}Re^{-2t/\tau}$$

$v(t)$ = voltage across the capacitor, at time t [V]
v_f = final voltage across the capacitor, steady-state voltage [V]
v_0 = initial voltage across the capacitor [V]
t = time [s]
τ = the time constant, RC [seconds]
C = capacitance [F]
Natural log: $\ln x = b \Leftrightarrow e^b = x$

</div>

<div style="border:1px solid">

C PARALLEL PLATE CAPACITANCE

$$C = \frac{\varepsilon A}{h} \qquad C_{\text{per unit length}} = \frac{\varepsilon A}{lh} = \frac{\varepsilon wl}{lh} = \frac{\varepsilon w}{h}$$

ε = permittivity of the material [F/cm]
A = area of one of the capacitor plates [cm^2]
h = plate separation [cm]
w = plate width [cm]
l = plate length [cm]
C = capacitance [F]
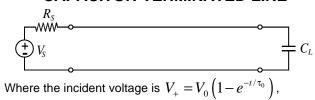
</div>

<div style="border:1px solid">

CAPACITOR-TERMINATED LINE

R_S

$+ V_S$ C_L

Where the incident voltage is $V_+ = V_0\left(1 - e^{-t/\tau_0}\right)$,

$$V_{cap} = V_+ + V_- = V_0\left(2 + \frac{2\tau_1}{\tau_0 - \tau_1}e^{-t/\tau_1} - \frac{2\tau_0}{\tau_0 - \tau_1}e^{-t/\tau_0}\right)$$

V_0 = final voltage across the capacitor [V]
t = time [s]
τ_0 = time constant of the incident wave, RC [s]
τ_1 = time constant effect due to the load, $Z_0 C_L$ [s]
C = capacitance [F]

</div>

<div style="border:1px solid">

SMITH CHART

First **normalize the load impedance** by dividing by the characteristic impedance, and find this point on the chart.

When working in terms of **reactance** X, an inductive load will be located on the top half of the chart, a capacitive load on the bottom half. It's the other way around when working in terms of **susceptance** B [Siemens].

Draw a straight line from the center of the chart through the normalized load impedance point to the edge of the chart.

Anchor a compass at the center of the chart and draw a circle through the normalized load impedance point. Points along this circle represent the normalized impedance at various points along the transmission line. Clockwise movement along the circle represents movement from the load toward the source with one full revolution representing 1/2 wavelength as marked on the outer circle. The two points where the circle intersects the horizontal axis are the voltage maxima (right) and the voltage minima (left).

The point opposite the impedance (180° around the circle) is the **admittance** Y [Siemens]. The reason admittance (or susceptibility) is useful is because admittances in parallel are simply added. (**Admittance** is the reciprocal of impedance; **susceptance** is the reciprocal of reactance.)

$$\boxed{\Gamma(z) = \Gamma_L e^{j2\beta z}}$$

$$e^{j2\beta z} = 1\angle 2\beta z$$

$$\mathbf{G}(z) = \frac{\mathbf{Z}(z) - 1}{\mathbf{Z}(z) + 1}$$

$$\mathbf{Z}_L = \frac{\Gamma_L - 1}{\Gamma_L + 1} \qquad \boxed{\mathbf{Z} = \frac{Z_L}{Z_0}}$$

z = distance from load [m]
$j = \sqrt{-1}$
ρ = magnitude of the reflection coefficient
β = phase constant
Γ = reflection coefficient
Z = normalized impedance [Ω]
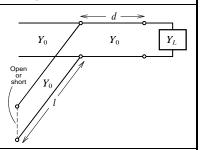
</div>
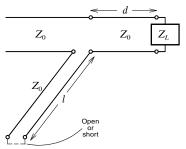
SINGLE-STUB TUNING

The basic idea is to connect a line stub in parallel (shunt) or series a distance d from the load so that the imaginary part of the load impedance will be canceled.

Shunt-stub: Select d so that the admittance Y looking toward the load from a distance d is of the form $Y_0 + jB$. Then the stub susceptance is chosen as $-jB$, resulting in a matched condition.
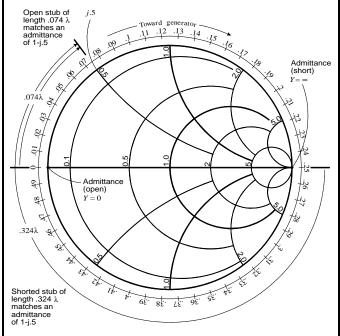
Series-stub: Select d so that the admittance Z looking toward the load from a distance d is of the form $Z_0 + jX$. Then the stub susceptance is chosen as $-jX$, resulting in a matched condition.

FINDING A STUB LENGTH

Example: Find the lengths of open and shorted shunt stubs to match an admittance of $1\text{-}j0.5$. The admittance of an open shunt (zero length) is $Y=0$; this point is located at the left end of the Smith Chart x-axis. We proceed clockwise around the Smith chart, i.e. away from the end of the stub, to the $+j0.5$ arc (the value needed to match $-j0.5$). The difference in the starting point and the end point on the wavelength scale is the length of the stub in wavelengths. The length of a shorted-type stub is found in the same manner but with the starting point at $Y=\infty$.

In this example, all values were in units of admittance. If we were interested in finding a stub length for a series stub problem, the units would be in impedance. The problem would be worked in exactly the same way. Of course in impedance, an open shunt (zero length) would have the value $Z=\infty$, representing a point at the right end of the x-axis.

LINE IMPEDANCE [Ω]

The impedance seen at the source end of a lossless transmission line:

$$Z_{in} = Z_0 \frac{1+\rho}{1-\rho} = Z_0 \frac{Z_L + jZ_0 \tan(\beta l)}{Z_0 + jZ_L \tan(\beta l)}$$

For a lossy transmission line:

$$Z_{in} = Z_0 \frac{Z_L + Z_0 \tanh(\gamma l)}{Z_0 + Z_L \tanh(\gamma l)}$$

Line impedance is periodic with spatial period $\lambda/2$.

$Z_0 = \sqrt{L/C}$, the characteristic impedance of the line. [Ω]

ρ = the reflection coefficient

Z_L = the load impedance [Ω]

$\beta = 2\pi/\lambda$, phase constant

$\gamma = \alpha + j\beta$, complex propagation constant

$\boldsymbol{\lambda}$ WAVELENGTH [cm]

The physical distance that a traveling wave moves during one period of its periodic cycle.

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{k} = \frac{v_p}{f}$$

$\beta = \omega\sqrt{LC}$ = $2\pi/\lambda$, phase constant
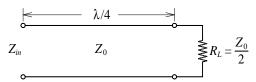
$k = \omega\sqrt{\mu\varepsilon}$ = $2\pi/\lambda$, wave number

v_p = velocity of propagation [m/s] see p 2.

f = frequency [Hz]

$\boldsymbol{\lambda}$/4 QUARTER-WAVE SECTION

A quarter-wave section of transmission line has the effect of inverting the _normalized_ impedance of the load.

To find Z_{in}, we can normalize the load (by dividing by the characteristic impedance), invert the result, and "unnormalize" this value by multiplying by the characteristic impedance.

In this case, the normalized load is $\dfrac{Z_0}{2} \div Z_0 = \dfrac{1}{2}$

so the normalized input impedance is $\left(\dfrac{1}{2}\right)^{-1} = 2$

and the actual input impedance is $Z_{in} = 2Z_0$

$\boldsymbol{\gamma}$ COMPLEX PROPAGATION CONSTANT

The propagation constant for lossy lines, taking into account the resistance along the line as well as the resistive path between the conductors.

$$\gamma = \alpha + j\beta = \sqrt{ZY} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$\alpha = \sqrt{RG}$ **attenuation constant**, the real part of the complex propagation constant, describes the loss

$\beta = 2\pi/\lambda$, **phase constant**, the complex part of the complex propagation constant

Z = **series impedance** (complex, inductive) per unit length [Ω/cm]

Y = **shunt admittance** (complex, capacitive) per unit length [Ω^{-1}/cm]

R = the resistance per unit length along the transmission line [Ω/cm]

G = the conductance between conductors per unit length [Ω^{-1}/cm]

L = inductance per unit length [H/cm]

C = capacitance per unit length [F/cm]

MODULATED WAVE

Suppose we have a disturbance composed of two frequencies:

$$\sin\left[(\omega_0 - \delta\omega)t - (\beta_0 - \delta\beta)z\right]$$

$$\text{and} \quad \sin\left[(\omega_0 + \delta\omega)t - (\beta_0 + \delta\beta)z\right]$$

where ω_0 is the average frequency and β_0 is the average phase.

Using the identity $2\cos\left(\dfrac{A-B}{2}\right)\sin\left(\dfrac{A+B}{2}\right) = \sin A + \sin B$

The combination (sum) of these two waves is

$$2\underbrace{\cos\left(\delta\omega t - \delta\beta z\right)}_{\text{envelope}}\underbrace{\sin\left(\omega_0 t - \beta_0 z\right)}_{\text{carrier}}$$

The envelope moves at the group velocity, see p 7.

δ = "the difference in"…
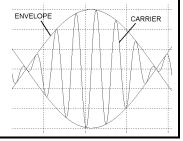
ω_0 = carrier frequency [radians/second]

ω = modulating frequency [radians/second]

β_0 = carrier frequency phase constant

β = phase constant

So the sum of two waves will be a modulated wave having a **carrier** frequency equal to the average frequency of the two waves, and an **envelope** with a frequency equal to half the difference between the two original wave frequencies.

ENVELOPE   CARRIER

v_g GROUP VELOCITY   [cm/s]

The velocity at which the envelope of a modulated wave moves.

$$v_g = \frac{\delta\omega}{\delta\beta} = \frac{1}{\sqrt{LC_P}}\sqrt{1 - \frac{\omega_c^2}{\omega^2}} \quad \text{where:}$$

L = inductance per unit length [H/cm]

C_P = capacitance per unit length [F/cm]

ε = permittivity of the material [F/cm]

μ = permeability of the material, dielectric constant [H/cm]

ω_c = carrier frequency [radians/second]

ω = modulating frequency [radians/second]
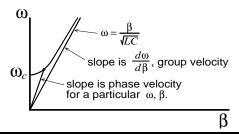
β = phase constant

Also, since β may be given as a function of ω, remember

$$v_g = \left(\frac{d\beta}{d\omega}\right)^{-1}$$

OMEGA - BETA GRAPH

This representation is commonly used for modulated waves.

$\omega = \frac{\beta}{\sqrt{LC}}$

slope is $\frac{d\omega}{d\beta}$, group velocity

slope is phase velocity for a particular ω, β.

d SKIN DEPTH   [cm]

The depth into a material at which a wave is attenuated by 1/e (about 36.8%) of its original intensity. This isn't the same δ that appears in the *loss tangent*, tan δ.

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega\mu\sigma}} \quad \text{where:}$$

$\alpha = \sqrt{RG}$ **attenuation constant**, the real part of the complex propagation constant, describes loss

μ = permeability of the material, dielectric constant [H/cm]

ω = frequency [radians/second]

σ = (*sigma*) conductivity [Siemens/meter] see p12.

Skin Depths of Selected Materials			
	60 Hz	**1 MHz**	**1 GHz**
silver	8.27 mm	0.064 mm	0.0020 mm
copper	8.53 mm	0.066 mm	0.0021 mm
gold	10.14 mm	0.079 mm	0.0025 mm
aluminum	10.92 mm	0.084 mm	0.0027 mm
iron	0.65 mm	0.005 mm	0.00016 mm

MAXWELL'S EQUATIONS

Maxwell's equations govern the principles of guiding and propagation of electromagnetic energy and provide the foundations of all electromagnetic phenomena and their applications. The time-harmonic expressions can be used only when the wave is sinusoidal.

	STANDARD FORM (Time Domain)	**TIME-HARMONIC** (Frequency Domain)
Faraday's Law	$\nabla \times \vec{\mathscr{E}} = -\dfrac{\partial \vec{\mathscr{B}}}{\partial t}$	$\nabla \times \vec{E} = -j\omega \vec{B}$
Ampere's Law*	$\nabla \times \vec{\mathscr{H}} = \vec{\mathscr{J}} + \dfrac{\partial \vec{\mathscr{D}}}{\partial t}$	$\nabla \times \vec{H} = j\omega \vec{D} + \vec{J}$
Gauss' Law	$\nabla \cdot \vec{\mathscr{D}} = \rho_v$	$\nabla \cdot \vec{D} = \rho_v$
no name law	$\nabla \cdot \vec{\mathscr{B}} = 0$	$\nabla \cdot \vec{B} = 0$

$\mathscr{E}$ = electric field [V/m]
$\mathscr{B}$ = magnetic flux density [W/m^2 or T] $\mathscr{B} = \mu_0 \mathscr{H}$
t = time [s]
$\mathscr{D}$ = electric flux density [C/m^2] $\mathscr{D} = \varepsilon_0 \mathscr{E}$
ρ = volume charge density [C/m^3]
$\mathscr{H}$ = magnetic field intensity [A/m]
$\mathscr{J}$ = current density [A/m^2]
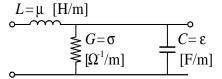
*Maxwell added the $\dfrac{\partial \vec{\mathscr{D}}}{\partial t}$ term to Ampere's Law.

ELECTROMAGNETIC WAVES

MODELING MAXWELL'S EQUATIONS

This is a model of a wave, analogous to a transmission line model.

L = inductance per unit length [H/cm]
μ = permeability of the material, dielectric constant [H/cm]
G = the conductance per unit length [Ω^{-1}/cm]
σ = (*sigma*) conductivity [Siemens/meter]
C = capacitance per unit length [F/cm]
ε = permittivity of the material [F/cm]

propagation constant: $\gamma = \sqrt{(j\omega\mu)(j\omega\varepsilon + \sigma)}$

LOW FREQUENCY

At low frequencies, more materials behave as **conductors**. A wave is considered low frequency when

$$\boxed{\omega \ll \frac{\sigma}{\varepsilon}}$$ $\dfrac{\sigma}{\varepsilon}$ is the dielectric relaxation frequency

$$\boxed{\eta = \frac{1}{\sigma\delta}(1 + j)}$$ intrinsic wave impedance, see p 12.

What happens to the complex propagation constant at low frequency? From the wave model above, gamma is

$$\gamma = \sqrt{(j\omega\mu)(j\omega\varepsilon + \sigma)} = \sqrt{j\omega\mu\sigma}\sqrt{1 + \frac{j\omega\varepsilon}{\sigma}}$$

Since both ω and ε/σ are small

$$\gamma = \sqrt{j\omega\mu\sigma}\left(1 + \frac{1}{2}j\omega\frac{\varepsilon}{\sigma}\right) = \sqrt{j\omega\mu\sigma}\,(1)$$

Since $\sqrt{j} = \frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}$

$$\gamma = \sqrt{\omega\mu\sigma}\left(\frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}\right) = \sqrt{\frac{\omega\mu\sigma}{2}} + j\sqrt{\frac{\omega\mu\sigma}{2}}$$

So that, with $\gamma = \alpha + j\beta$

we get $\boxed{\alpha = \sqrt{\dfrac{\omega\mu\sigma}{2}}}$, $\boxed{\beta = \sqrt{\dfrac{\omega\mu\sigma}{2}}}$ or $\boxed{\gamma = \dfrac{1}{\delta}(1 + j)}$

α = **attenuation constant**, the real part of the complex propagation constant, describes the loss
β = **phase constant**, the complex part of the complex propagation constant
σ = (*sigma*) conductivity [Siemens/cm]
δ = skin depth [cm]
So the wave is attenuating at the same rate that it is propagating.

HIGH FREQUENCY

At high frequencies, more materials behave as **dielectrics**, i.e. copper is a dielectric in the gamma ray range. A wave is considered high frequency when

$$\omega \gg \frac{\sigma}{\varepsilon}$$ $\dfrac{\sigma}{\varepsilon}$ is the dielectric relaxation frequency

$$\eta = \sqrt{\frac{\mu}{\varepsilon}}$$ intrinsic wave impedance, see p 12.

What happens to the complex propagation constant at high frequency?

$$\gamma = \sqrt{(j\omega\mu)(j\omega\varepsilon + \sigma)} = \sqrt{j\omega\mu \, j\omega\varepsilon \left(1 + \frac{\sigma}{j\omega\varepsilon}\right)}$$

Since both $1/\omega$ and σ/ε are small

$$\gamma = j\omega\sqrt{\mu\varepsilon}\left(1 + \frac{1}{2}\frac{\sigma}{j\omega\varepsilon}\right) \quad \boxed{\gamma = \frac{\sigma}{2}\sqrt{\frac{\mu}{\varepsilon}} + j\omega\sqrt{\mu\varepsilon}}$$

With $\gamma = \alpha + j\beta$

we get $\boxed{\alpha = \dfrac{\sigma}{2}\sqrt{\dfrac{\mu}{\varepsilon}}}$, $\boxed{\beta = \omega\sqrt{\mu\varepsilon}}$,

tan **d** LOSS TANGENT

The loss tangent, a value between 0 and 1, is the loss coefficient of a wave after it has traveled one wavelength. This is the way data is usually presented in texts. This is not the same δ that is used for skin depth.

$$\boxed{\tan\delta = \frac{\sigma}{\omega\varepsilon}}$$

Graphical representation of loss tangent:

For a dielectric, $\tan\delta \ll 1$.

$$\alpha \approx \frac{1}{2}(\tan\delta)\beta = \frac{\pi}{\lambda}\tan\delta$$

Imag. (*I*)

$\omega\varepsilon$

δ

Re (*I*)

σ

$\omega\varepsilon$ is proportional to the amount of current going through the capacitance *C*.
σ is proportional to the amount current going through the conductance *G*.

TEM WAVES
Transverse Electromagnetic Waves

Electromagnetic waves that have single, orthogonal vector electric and magnetic field components (e.g., $\mathscr{E}_x$ and $\mathscr{H}_y$), both varying with a single coordinate of space (e.g., *z*), are known as *uniform plane waves* or *transverse electromagnetic (TEM) waves*. TEM calculations may be made using formulas from electrostatics; this is referred to as *quasi-static* solution.

Characteristics of TEM Waves

- The velocity of propagation (always in the *z* direction) is $v_p = 1/\sqrt{\mu\varepsilon}$, which is the speed of light in the material
- There is no electric or magnetic field in the direction of propagation. Since this means there is no voltage drop in the direction of propagation, it suggests that no current flows in that direction.
- The electric field is normal to the magnetic field
- The value of the electric field is η times that of the magnetic field at any instant.
- The direction of propagation is given by the direction of $\mathbf{E} \times \mathbf{H}$.
- The energy stored in the electric field per unit volume at any instant and any point is equal to the energy stored in the magnetic field.

TEM ASSUMPTIONS

Some assumptions are made for TEM waves.

$\mathscr{E}_z = 0$ $\qquad\qquad\qquad$ $\mathscr{H}_z = 0$

$\sigma = 0$ $\qquad\qquad\qquad$ time dependence $e^{j\omega t}$

WAVE ANALOGIES

Plane waves have many characteristics analogous to transmission line problems.

Transmission Lines	Plane Waves
Phase constant $$\beta = \omega\sqrt{LC} = \frac{\omega}{v_p} = \frac{2\pi}{\lambda}$$	Wave number $$k = \omega\sqrt{\mu\varepsilon} = \frac{\omega}{v_p} = \frac{2\pi}{\lambda}$$
Complex propagation const. $$\gamma = \alpha + j\beta$$ $$= \sqrt{(R + j\omega L)(G + j\omega C)}$$	Complex propagation constant $$\gamma = \sqrt{(j\omega\mu)(j\omega\varepsilon + \sigma)}$$
Velocity of propagation $$v_p = \frac{1}{\sqrt{LC}} = \frac{\omega}{\beta}$$	Phase velocity $$v_p = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{\omega}{k} = \omega\delta = c\frac{2\pi\delta}{\lambda}$$
Characteristic impedance $$Z_0 = \sqrt{\frac{L}{C}} = \frac{V_+}{I_+}$$	Intrinsic impedance $$\eta = \sqrt{\frac{\mu}{\varepsilon}} = \frac{E_{x+}}{H_{y+}}$$
Voltage $$V(z) = V_+ e^{-j\beta z} + V_- e^{j\beta z}$$	Electric Field $$E_x(z) = E_+ e^{-jkz} + E_- e^{jkz}$$
Current $$I(z) = \frac{1}{Z_0}\left[V_+ e^{-j\beta z} - V_- e^{j\beta z}\right]$$	Magnetic Field $$H_y(z) = \frac{1}{\eta}\left[E_+ e^{-jkz} - E_- e^{jkz}\right]$$
Line input impedance $$Z_{in} = Z_0\frac{Z_L + jZ_0\tan(\beta l)}{Z_0 + jZ_L\tan(\beta l)}$$ $$Z_{in} = Z_0\frac{Z_L + Z_0\tanh(\gamma l)}{Z_0 + Z_L\tanh(\gamma l)}$$	Wave input impedance $$\eta_{in} = \eta_0\frac{\eta_L + j\eta_0\tan(kl)}{\eta_0 + j\eta_L\tan(kl)}$$ $$\eta_{in} = \eta_0\frac{\eta_L + \eta_0\tanh(\gamma l)}{\eta_0 + \eta_L\tanh(\gamma l)}$$
Reflection coefficient $$\rho = \frac{Z_L - Z_0}{Z_L + Z_0}$$	Reflection coefficient $$\rho = \frac{\eta_L - \eta_0}{\eta_L + \eta_0}$$

k WAVE NUMBER [rad./cm]

The phase constant for the uniform plane wave; the change in phase per unit length. It can be considered a constant for the medium at a particular frequency.

$$k = \frac{\omega}{v} = \omega\sqrt{\mu\varepsilon} = \frac{2\pi}{\lambda}$$

k appears in the phasor forms of the uniform plane wave

$$E_x(z) = E_1 e^{-jkz} + E_2 e^{jkz} \text{, etc.}$$

k has also been used as in the "k of a dielectric" meaning ε_r.

$\mathbf{h}$ (eta) INTRINSIC WAVE IMPEDANCE [Ω]

The ratio of electric to magnetic field components. Can be considered a constant of the medium. For free space, $\eta = 376.73\Omega$. The units of η are in ohms.

$$\eta = \frac{E_{x+}}{H_{y+}} = -\frac{E_{y+}}{H_{x+}} \qquad -\eta = \frac{E_{x-}}{H_{y-}} = -\frac{E_{y-}}{H_{x-}}$$
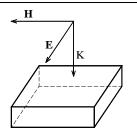
at low frequencies	at high frequencies
$$\eta = \frac{1}{\sigma\delta}(1 + j)$$	$$\eta = \sqrt{\frac{\mu}{\varepsilon}}$$
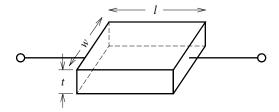
When an electromagnetic wave encounters a sheet of conductive material it sees an impedance. K is the direction of the wave, **H** is the magnetic component and **E** is the electrical field. $\mathbf{E} \times \mathbf{H}$ gives the direction of propagation K.

SHEET RESISTANCE   [Ω]

Consider a block of material with conductivity σ.

It's resistance is $R = \dfrac{l}{wt\sigma}$ Ω.

If the length is equal to the width, this reduces to
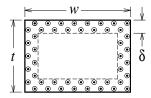
$$R = \frac{1}{t\sigma} \ \ \Omega.$$

And this is sheet resistance.

HIGH FREQUENCY RESISTANCE   [Ω]

When a conductor carries current at high frequency, the electric field penetrates the outer surface only about 1 skin depth so that current travels near the surface of the conductor. Since the entire cross-section is not utilized, this affects the resistance of the conductor.

Cross-section of a conductor showing current flow near the surface:

$$R \approx \frac{1}{\sigma\delta(\text{perimeter})} = \sqrt{\frac{\omega\mu_0}{2\sigma}} \ \frac{1}{2w+2t}$$

σ = (*sigma*) conductivity (5.8×10^5 S/cm for copper) [Siemens/meter]
ω = frequency [radians/second]
δ = skin depth [cm]
μ_0 = permeability of free space $\mu_0 = 4\pi\times10^{-9}$ [H/cm]
w = width of the conductor [cm]
t = thickness of the conductor [cm]
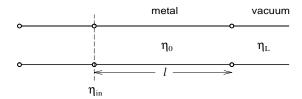
$\mathbf{h_{in}}$ WAVE INPUT IMPEDANCE   [Ω]

The impedance seen by a wave in a medium.

For example, the impedance of a metal sheet in a vacuum:

Note that a transmission line model is used here because it is <u>analogous</u> to a wave traveling in a medium. The "load" is the element most remote in the direction of propagation.

The input impedance is $\eta_{in} = \eta_0 \dfrac{\eta_L + \eta_0 \tanh(\gamma l)}{\eta_0 + \eta_L \tanh(\gamma l)}$ Ω.

In this example, *l* is the thickness of a metal sheet. If the metal thickness is much greater than the skin depth, then

$$\tanh(\lambda l) = \tanh\left[\frac{1}{\delta}(1+j)l\right] = \tanh\left[(\text{big number})(1+j)\right] \approx 1$$

If *l* is much less than the skin depth δ, then

$$\tanh(\lambda l) = \tanh\left[\frac{1}{\delta}(1+j)l\right] = \tanh\left[(\text{small number})(1+j)\right]$$

$$= (\text{same small number})(1+j) = \frac{l}{\delta}(1+j)$$

$\mathbf{m}$ MAGNETIC PERMEABILITY   [H/m]

The relative increase or decrease in the resultant magnetic field inside a material compared with the magnetizing field in which the given material is located. The product of the permeability constant and the relative permeability of the material.

$$\mu = \mu_0\mu_r \quad \text{where } \mu_0 = 4\pi\times10^{-7} \text{ H/m}$$

Relative Permeabilities of Selected Materials			
Air	1.00000037	Mercury	0.999968
Aluminum	1.000021	Nickel	600
Copper	0.9999833	Oxygen	1.000002
Gold	0.99996	Platinum	1.0003
Iron (99.96% pure)	280,000	Silver	0.9999736
Iron (motor grade)	5000	Titanium	1.00018
Lead	0.9999831	Tungsten	1.00008
Manganese	1.001	Water	0.9999912

e ELECTRIC PERMITTIVITY [F/m]

The property of a dielectric material that determines how much electrostatic energy can be stored per unit of volume when unit voltage is applied, also called the **dielectric constant**. The product of the constant of permittivity and the relative permittivity of a material.

$$\varepsilon = \varepsilon_0 \varepsilon_r \quad \text{where } \varepsilon_0 = 8.85\times10^{-14} \text{ F/cm}$$

---

e$_c$ COMPLEX PERMITTIVITY

$$\varepsilon_c = \varepsilon' - j\varepsilon'' \quad \text{where } \frac{\varepsilon''}{\varepsilon'} = \tan\delta_c$$

In general, both ε' and ε'' depend on frequency in complicated ways. ε' will typically have a constant maximum value at low frequencies, tapering off at higher frequencies with several peaks along the way. ε'' will typically have a peak at the frequency at which ε' begins to decline in magnitude as well as at frequencies where ε' has peaks, and will be zero at low frequencies and between peaks.

---

e$_r$ RELATIVE PERMITTIVITY

The permittivity of a material is the relative permittivity multiplied by the permittivity of free space

$$\varepsilon = \varepsilon_r \times \varepsilon_0$$

In old terminology, ε_r is called the "k of a dielectric". Glass (SiO$_2$) at ε_r = 4.5 is considered the division between low k and high k dielectrics.

Relative Permittivities of Selected Materials

Air (sea level)	1.0006	Polystyrene	2.6
Ammonia	22	Polyethylene	2.25
Bakelite	5	Rubber	2.2-4.1
Glass	4.5-10	Silicon	11.9
Ice	3.2	Soil, dry	2.5-3.5
Mica	5.4-6	Styrofoam	1.03
most metals	~1	Teflon	2.1
Plexiglass	3.4	Vacuum	1
Porcelain	5.7	Water, distilled	81
Paper	2-4	Water, seawater	72-80
Oil	2.1-2.3		

NOTE: Relative permittivity data is given for materials at **low or static frequency conditions**. The permittivity for most materials varies with frequency. The relative permittivities of most materials lie in the range of 1-25. At high frequencies, the permittivity of a material can be quite different (usually less), but will have resonant peaks.

---

s CONDUCTIVITY [S/m] or [1/(Ω·m)]

A measure of the ability of a material to conduct electricity, the higher the value the better the material conducts. The reciprocal is *resistivity*. Values for common materials vary over about 24 orders of magnitude. Conductivity may often be determined from skin depth or the loss tangent.

$$\sigma = \frac{n_c q_e^{\,2} \overline{l}}{m_e v_{th}} \text{ S/m} \quad \text{where}$$

n_c = density of conduction electrons (for copper this is 8.45×10^{28}) [m^{-3}]

q_e = electron charge? 1.602×10^{-23} [C]

$\overline{l}$ = $v_{th}t_c$ the product of the thermal speed and the mean free time between collisions of electrons, the average distance an electron travels between collisions [m]

m_e = the effective electron mass? [kg]

v_{th} = thermal speed, usually much larger than the drift velocity v_d. [m/s]

Conductivities of Selected Materials [1/(Ω·m)]

Aluminum	3.82×10^7	Mercury	1.04×10^6
Carbon	7.14×10^4	Nicrome	1.00×10^6
Copper (annealed)	5.80×10^7	Nickel	1.45×10^7
Copper (in class)	6.80×10^7	Seawater	4
Fresh water	~10^{-2}	Silicon	~4.35×10^{-4}
Germanium	~2.13	Silver	6.17×10^7
Glass	~10^{-12}	Sodium	2.17×10^7
Gold	4.10×10^7	Stainless steel	1.11×10^6
Iron	1.03×10^7	Tin	8.77×10^6
Lead	4.57×10	Titanium	2.09×10^6
		Zinc	1.67×10^7

P POWER [W]
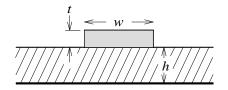
Power is the time rate of change of energy.

Power reflected at a discontinuity: $\% \text{ power} = |\rho|^2 \times 100$

Power transmitted at a discontinuity: $\% \text{ power} = \left(1 - |\rho|^2\right) \times 100$

MICROSTRIP CONDUCTORS

How fast does a wave travel in a microstrip? The question is complicated by the fact that the dielectric on one side of the strip may be different from the dielectric on the other side and a wave may travel at different speeds in different dielectrics. The solution is to find an **effective relative permittivity** $\varepsilon_{r\,\mathrm{eff}}$ for the combination.

Some Microstrip Relations

$$Z_0^{\mathrm{air}} = Z_0\sqrt{\varepsilon_{r\mathrm{eff}}} \qquad\qquad C^{\mathrm{air}}Z_0^{\mathrm{air}} = \sqrt{\varepsilon_0\mu_0}$$

$$L = Z_0^{\mathrm{air}}\sqrt{\varepsilon_0\mu_0} = C^{\mathrm{total}}\left(Z_0\right)^2 \qquad LC^{\mathrm{air}} = \varepsilon_0\mu_0$$

$$Z_0 = \sqrt{\frac{L}{C^{\mathrm{total}}}} \qquad\qquad Z_0^{\mathrm{air}} = \sqrt{\frac{L}{C^{\mathrm{air}}}}$$

$$\gamma = j\beta = j\omega\sqrt{\varepsilon_0\mu_0}\sqrt{\varepsilon_{r\mathrm{eff}}} \qquad \varepsilon_{r\mathrm{eff}} = \frac{C^{\mathrm{total}}}{C^{\mathrm{air}}}$$

$$v_p = \frac{1}{\sqrt{\varepsilon_0\mu_0\varepsilon_{r\mathrm{eff}}}} = \frac{1}{\sqrt{LC^{\mathrm{total}}}}$$

It's difficult to get more than 200Ω for Z_0 in a microstrip.

Microstrip Approximations
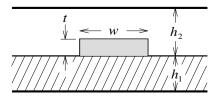
$$\varepsilon_{r\mathrm{eff}} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2\sqrt{1+12h/w}}$$

$$Z_0 = \begin{cases} \dfrac{60}{\sqrt{\varepsilon_{r\mathrm{eff}}}}\ln\left[\dfrac{8h}{w}+\dfrac{w}{4h}\right], & \text{for } \dfrac{w}{h}\le 1 \\[3ex] \dfrac{120\pi}{\sqrt{\varepsilon_{r\mathrm{eff}}}\left[\dfrac{w}{h}+1.393+0.667\ln\left(\dfrac{w}{h}+1.444\right)\right]}, & \text{for } \dfrac{w}{h}>1 \end{cases}$$

$$\frac{w}{h} = \begin{cases} \dfrac{8e^A}{e^{2A}-2}, & \dfrac{w}{h}<2 \\[3ex] \dfrac{2}{\pi}\left\{B-1-\ln(2B-1)+\dfrac{\varepsilon_r-1}{2\varepsilon_r}\left[\ln(B-1)+0.39-\dfrac{0.61}{\varepsilon_r}\right]\right\}, & \dfrac{w}{h}>2 \end{cases}$$

where $A = \dfrac{Z_0}{60}\sqrt{\dfrac{\varepsilon_r+1}{2}}+\dfrac{\varepsilon_r-1}{\varepsilon_r+1}\left(0.23+\dfrac{0.11}{\varepsilon_r}\right)$, $B = \dfrac{377\pi}{2Z_0\sqrt{\varepsilon_r}}$

STRIPLINE CONDUCTOR

Also called *shielded microstrip*. The effective relative permittivity is used in calculations.

assuming $w \ge 10h$, $\qquad \varepsilon_{r\mathrm{eff}} = \dfrac{\varepsilon_{r1}h_1+\varepsilon_{r2}h_2}{h_1+h_2}$ where

ε_{r1} = the relative permittivity of the dielectric of thickness h_1.
ε_{r2} = the relative permittivity of the dielectric of thickness h_2.

COPPER CLADDING

The thickness of copper on a circuit board is measured in ounces. 1-ounce cladding means that 1 square foot of the copper weighs 1 ounce. 1-ounce copper is 0.0014" or 35.6 μm thick.

$\mathbf{a}_d$ DIELECTRIC LOSS FACTOR [dB/cm]

$$\alpha_d = 8.68\frac{\beta_0\varepsilon_r\left(\varepsilon_{r\mathrm{eff}}-1\right)}{2\sqrt{\varepsilon_{r\mathrm{eff}}}\left(\varepsilon_r-1\right)}\tan\delta$$

$\mathbf{a}_c$ CONDUCTOR LOSS FACTOR [dB/cm]

$$\alpha_c = 8.68\frac{R}{2Z_0}, \qquad R = \frac{1}{\sigma\delta(\mathrm{perimeter})} = \sqrt{\frac{\omega\mu_0}{2\sigma}}\frac{1}{(\mathrm{perimeter})}$$

WHEELER'S EQUATION
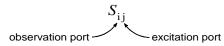
Another approximation for microstrip calculations is Wheeler's equation.

$$Z_0 = \frac{42.4}{\sqrt{1+\varepsilon_r}} \ln\left(1 + \left\{\frac{4h}{w'}\left[\frac{14+\frac{8}{\varepsilon_r}}{11}\times\frac{4h}{w'} + \sqrt{\left(\frac{14+\frac{8}{\varepsilon_r}}{11}\times\frac{4h}{w'}\right)^2 + \pi^2\frac{1+\frac{1}{\varepsilon_r}}{2}}\right]\right\}\right)$$

where
$$w' = \frac{8h\sqrt{\frac{7+\frac{4}{\varepsilon_r}}{11}\left[\exp\left(\frac{Z_0}{42.4}\sqrt{\varepsilon_r+1}\right)-1\right]+\frac{1+\frac{1}{\varepsilon_r}}{0.81}}}{\exp\left(\frac{Z_0}{42.4}\sqrt{\varepsilon_r+1}\right)-1}$$

NETWORK THEORY

S_{ij} SCATTERING PARAMETER

$$S_{ij}$$

observation port — excitation port

A scattering parameter, represented by S_{ij}, is a dimensionless value representing the fraction of wave amplitude transmitted from port j into port i, provided that all other ports are terminated with matched loads and only port j is receiving a signal. Under these same conditions, S_{ii} is the reflection coefficient at port i.

To experimentally determine the scattering parameters, attach an impedance-matched generator to one of the ports (**excitation port**), attach impedance-matched loads to the remaining ports, and observe the signal received at each of the ports (**observation ports**). The fractional amounts of signal amplitude received at each port i will make up one column j of the **scattering matrix**. Repeating the process for each column would require n^2 measurements to determine the scattering matrix for an n-port network.

S_{ij} SCATTERING MATRIX

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{bmatrix}$$

The scattering matrix is an $n \times n$ matrix composed of scattering parameters that describes an n-port network.

The elements of the diagonal of the scattering matrix are reflection coefficients of each port. The elements of the off-diagonal are transmission coefficients, under the conditions outlined in "SCATTERING PARAMETER".

If the network is **internally matched** or **self-matched**, then $S_{11} = S_{22} = \cdots = S_{NN} = 0$, that is, the diagonal is all zeros.

The sum of the squares of each column of a scattering matrix is equal to one, provided the network is lossless.

a_n, b_n INCIDENT/REFLECTED WAVE AMPLITUDES

The parameters a_n and b_n describe the incident and reflected waves respectively at each port n. These parameters are used for power and scattering matrix calculations.

The amplitude of the wave incident to port n is equal to the amplitude of the incident voltage at the port divided by the square root of the port impedance.	$a_n = \dfrac{V_n^+}{\sqrt{Z_{0n}}}$
Amplitude of the wave reflected at port n is equal to the amplitude of the reflected voltage at the port divided by the square root of the port impedance.	$b_n = \dfrac{V_n^-}{\sqrt{Z_{0n}}}$
The scattering parameter is equal to the wave amplitude output at port i divided by the wave amplitude input at port j provided the only source is a matched source at port j and all other ports are connected to matched loads.	$S_{ij} = \dfrac{b_i}{a_j}$
The relationship between the S-parameters and the a- and b-parameters can be written in matrix form where $\mathbf{S}$ is the scattering matrix and $\mathbf{a}$ and $\mathbf{b}$ are column vectors.	$\mathbf{b} = \mathbf{Sa}$
Power flow into any port is shown as a function of a- and b-parameters.	$P = \dfrac{1}{2}\left(|a|^2 - |b|^2\right)$
The ratio of the input power at port j to the output power at port I can be written as a function of a- and b-parameters or the S-parameter.	$\dfrac{P_{inj}}{P_{outi}} = \dfrac{|a_j|^2}{|b_j|^2} = \dfrac{1}{|S_{ij}|^2}$

RECIPROCITY

A network is reciprocal when $S_{ij} = S_{ji}$ in the scattering matrix, i.e. the matrix is symmetric across the diagonal. Also, $Z_{ij} = Z_{ji}$ and $Y_{ij} = Y_{ji}$. Networks constructed of "normal materials" exhibit reciprocity.

Reciprocity Theorem:

$$\oint_S \vec{E}_a \times \vec{H}_b \cdot ds = \oint_S \vec{E}_b \times \vec{H}_a \cdot ds$$

E_a and H_b are fields from two different sources.

LOSSLESS NETWORK

A network is lossless when

$$\underline{S}\ \underline{S}^\dagger = /$$

† means to take the complex conjugate and transpose the matrix. If the network is reciprocal, then the transpose is the same as the original matrix.

/ = a unitary matrix. A **unitary matrix** has the properties:
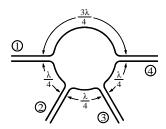
$$\sum_{k=1}^{N} S_{ki} S_{ki}^* = 1 \qquad\qquad \sum_{k=1}^{N} S_{ki} S_{kj}^* = 0$$

In other words, a column of a unitary matrix multiplied by its complex conjugate equals one, and a column of a unitary matrix multiplied by the complex conjugate of a different column equals zero.

RAT RACE OR HYBRID RING NETWORK
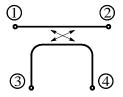
The rat race or hybrid ring network is lossless, reciprocal, and internally matched.

The signal splits upon entering the network and half travels around each side. A signal entering at port 1 and exiting at port 4 travels ¾ of a wavelength along each side, so the signals are in phase and additive. From port 1 to port 3 the signal travels one wavelength along one side and ½ wavelength along the other, arriving a port 3 out of phase and thus canceling. From port 1 to port 2 the paths are ¼ and 5/4 wavelengths respectively, thus they are in phase and additive.

DIRECTIONAL COUPLER

The directional coupler is a 4-port network similar to the rat race. It can be used to measure reflected and transmitted power to an antenna.
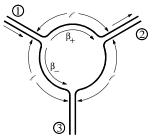
An input at one port is divided between two of the remaining ports. The **coupling factor**, measured in dB, describes the division of signal strength at the two ports. For example if the coupler has a coupling factor of –10 dB, then a signal input at port 1 would appear at port 4 attenuated by 10 dB with the majority of the signal passing to port 2. In other words, 90% of the signal would appear at port 2 and 10% at port 4. (-10 dB means "**10 dB down**" or 0.1 power, -6 dB means 0.25 power, and –3 dB means 0.5 power.) A reflection from port 2 would appear at port 3 attenuated by the same amount. Meters attached to ports 3 and 4 could be used to measure reflected and transmitted power for a system with a transmitter connected to port 1 and an antenna at port 2. The **directivity** of a coupler is a measurement of how well the coupler transfers the signal to the appropriate output without reflection due to the coupler itself; the directivity approaches infinity for a perfect coupler.

$\text{directivity} = 10\log\left(p_3 / p_1\right)$, where the source is at port 1 and the load is at port 2.

The directional coupler is **lossless** and **reciprocal**. The scattering matrix looks like this. In a real coupler, the off-diagonal zeros would be near zero due to leakage.

$$\begin{bmatrix} 0 & p & 0 & -q \\ p & 0 & q & 0 \\ 0 & q & 0 & p \\ -q & 0 & p & 0 \end{bmatrix}$$

CIRCULATOR

The circulator is a 3-port network that can be used to prevent reflection at the antenna from returning to the source.

Port 3 is terminated internally by a matched load. With a source at 1 and a load at 2, any power reflected at the load is absorbed by the load resistance at port 3. A 3-port network cannot be both lossless and reciprocal, so the circulator is <u>not</u> reciprocal.

Schematically, the circulator may be depicted like this:

The circulator is **lossless** but is <u>not</u> **reciprocal**. The scattering matrix looks like this:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

MAXWELL'S EQUATIONS, TIME HARMONIC FORM

$$\nabla \times \bar{\mathscr{E}} = -j\omega\mu\mathbf{H} \qquad \text{"curl on E"}$$

$$\nabla \times \bar{\mathscr{H}} = -j\omega\mu\mathbf{E} \qquad \text{"curl on H"}$$

$$\bar{\mathscr{E}} = \left[E_x(x,y)\hat{\mathbf{x}} + E_y(x,y)\hat{\mathbf{y}} + E_z(x,y)\hat{\mathbf{z}} \right] e^{j\omega t - \gamma z}$$

$$\bar{\mathscr{H}} = \left[H_x(x,y)\hat{\mathbf{x}} + H_y(x,y)\hat{\mathbf{y}} + H_z(x,y)\hat{\mathbf{z}} \right] e^{j\omega t - \gamma z}$$

From the curl equations we can derive:

(1) $\dfrac{\partial E_z}{\partial y} + \gamma E_y = -j\omega\mu H_x$ (4) $\dfrac{\partial H_z}{\partial y} + \gamma H_y = j\omega\varepsilon E_x$

(2) $-\dfrac{\partial E_z}{\partial x} - \gamma E_x = -j\omega\mu H_y$ (5) $-\dfrac{\partial H_z}{\partial x} - \gamma H_x = j\omega\varepsilon E_y$

(3) $\dfrac{\partial E_y}{\partial x} - \dfrac{\partial E_x}{\partial y} = -j\omega\mu H_z$ (6) $\dfrac{\partial H_y}{\partial x} - \dfrac{\partial H_x}{\partial y} = j\omega\varepsilon E_z$

From the above equations we can obtain:

(1) & (5) $H_x = \dfrac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left(j\omega\varepsilon\dfrac{\partial E_z}{\partial y} - \gamma\dfrac{\partial H_z}{\partial x} \right)$

(2) & (4) $H_y = \dfrac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left(j\omega\varepsilon\dfrac{\partial E_z}{\partial x} - \gamma\dfrac{\partial H_z}{\partial y} \right)$

(2) & (4) $E_x = -\dfrac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left(-\gamma\dfrac{\partial E_z}{\partial x} + j\omega\mu\dfrac{\partial H_z}{\partial y} \right)$

(1) & (5) $E_y = -\dfrac{1}{\gamma^2 + \omega^2\mu\varepsilon}\left(-\gamma\dfrac{\partial E_z}{\partial y} + j\omega\mu\dfrac{\partial H_z}{\partial x} \right)$

This makes it look like if E_z and H_z are zero, then H_x, H_y, E_x, and E_y are all zero. But since $\infty \times 0 \neq 0$, we could have non-zero result for the TEM wave if $\gamma^2 = -\omega^2\mu\varepsilon \Rightarrow \gamma = j\omega\sqrt{\mu\varepsilon}$. This should look familiar.

WAVE EQUATIONS

From Maxwell's equations and a vector identity on curl, we can get the following wave equations:

$$\nabla^2 \overline{\mathscr{E}} = -\omega^2 \mu \varepsilon \overline{\mathscr{E}} \qquad \text{"del squared on E"}$$

$$\nabla^2 \overline{\mathscr{H}} = \omega^2 \mu \varepsilon \overline{\mathscr{H}} \qquad \text{"del squared on H"}$$

The z part or "del squared on E_z" is:

$$\nabla^2 E_z = \frac{\gamma^2 E_z}{\partial x^2} + \frac{\gamma^2 E_z}{\partial y^2} + \frac{\gamma^2 E_z}{\partial z^2} = -\omega^2 \mu \varepsilon E_z$$

Using the separation of variables, we can let:

$$E_z = X(x) \cdot Y(y) \cdot Z(z)$$

We substitute this into the previous equation and divide by $X \cdot Y \cdot Z$ to get:

$$\underbrace{\frac{1}{X}\frac{d^2 X}{dx^2}}_{-k_x^2} + \underbrace{\frac{1}{Y}\frac{d^2 Y}{dy^2}}_{-k_y^2} + \underbrace{\frac{1}{Z}\frac{d^2 Z}{dz^2}}_{-k_z^2} = \underbrace{-\omega^2 \mu \varepsilon}_{\text{a constant}}$$

Since X, Y, and Z are independent variables, the only way the sum of these 3 expressions can equal a constant is if all 3 expressions are constants.

So we are letting $\dfrac{1}{Z}\dfrac{d^2 Z}{dz^2} = -k_z^2 \implies \dfrac{d^2 Z}{dz^2} = -Z k_z^2$

A solution could be $Z = e^{-\gamma z}$

so that $\gamma^2 e^{-\gamma z} = -k_z^2 e^{-\gamma z}$ and $-k_z^2 = \gamma^2$

Solutions for X and Y are found

$$\frac{1}{X}\frac{d^2 X}{dx^2} = -k_x^2 \implies X = A\sin(k_x x) + B\cos(k_x x)$$

$$\frac{1}{Y}\frac{d^2 Y}{dy^2} = -k_y^2 \implies Y = C\sin(k_y y) + D\cos(k_y y)$$

giving us the general solution $k_x^2 + k_y^2 - \gamma^2 = \omega^2 \mu \varepsilon$

For a particular solution we need to specify initial conditions and boundary conditions. For some reason, initial conditions are not an issue. The unknowns are k_x, k_y, A, B, C, D. The boundary conditions are

$$E_{\tan} = 0 \qquad\qquad \frac{\partial H_{\tan}}{\partial n} = 0$$
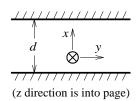
$E_{\tan}$ = the electric field tangential to a conducting surface
$H_{\tan}$ = the magnetic field tangential to a conducting surface
n = I don't know

TM, TE WAVES IN PARALLEL PLATES

TM, or transverse magnetic, means that magnetic waves are confined to the *transverse plane*. Similarly, TE (transverse electric) means that electrical waves are confined to the transverse plane.

(z direction is into page)

Transverse plane means the plane that is transverse to (perpendicular to) the direction of propagation. The direction of propagation is taken to be in the z direction, so the transverse plane is the x-y plane. So for a TM wave, there is no H_z component (magnetic component in the z direction) but there is an E_z component.

$$\boxed{E_z = A\sin(k_x x)e^{-\gamma z}}$$

A = amplitude [V]

$k_x = \dfrac{m\pi}{d}$ The magnetic field must be zero at the plate

boundaries. This value provides that characteristic. [cm^{-1}]

x = position; perpendicular distance from one plate. [cm]
d = plate separation [cm]
γ = propagation constant
z = position along the direction of propagation [cm]
m = mode number; an integer greater than or equal to 1
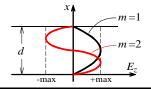
$$\boxed{\gamma = \sqrt{-\omega^2 \mu \varepsilon + (kx)^2}}$$

Notice than when $(kx)^2 \ge \omega^2 \mu \varepsilon$, the quantity under the square root sign will be positive and γ will be purely real. In this circumstance, the wave is said to be evanescent. The wavelength goes to infinity; there is no oscillation or propagation. On the other hand, when $(kx)^2 < \omega^2 \mu \varepsilon$, γ is purely imaginary.

The magnitude of E_z is related to its position between the plates and the mode number m. Note that for $m = 2$ that $d = \lambda$.

GENERAL MATHEMATICAL
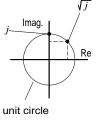
COMPLEX TO POLAR NOTATION

j in polar notation:

$$j = e^{j\frac{\pi}{2}}$$

So we can find the square root of j:

$$\sqrt{j} = \sqrt{e^{j\frac{\pi}{2}}} = e^{j\frac{\pi}{4}} = \frac{1}{\sqrt{2}} + j\frac{1}{\sqrt{2}}$$

Imag.

$\sqrt{j}$

j

Re

unit circle

dBm   DECIBELS RELATIVE TO 1 mW

The decibel expression for power. The logarithmic nature of decibel units translates the multiplication and division associated with gains and losses into addition and subtraction.

   0 dBm = 1 mW
  20 dBm = 100 mW
 -20 dBm = 0.01 mW

$$P(\text{dBm}) = 10\log\left[P(\text{mW})\right]$$
$$P(\text{mW}) = 10^{P(\text{dBm})/10}$$

PHASOR NOTATION

To express a derivative in phasor notation, replace $\dfrac{\partial}{\partial t}$ with $j\omega$. For example, the

Telegrapher's equation $\dfrac{\partial V}{\partial z} = -L\dfrac{\partial I}{\partial t}$

becomes $\dfrac{\partial V}{\partial z} = -Lj\omega I$.

Ñ NABLA, DEL OR GRAD OPERATOR

Compare the ∇ operation to taking the time derivative. Where $\partial/\partial t$ means to take the derivative with respect to time and introduces a s^{-1} component to the units of the result, the ∇ operation means to take the derivative with respect to distance (in 3 dimensions) and introduces a m^{-1} component to the units of the result. ∇ terms may be called space derivatives and an equation which contains the ∇ operator may be called a vector differential equation. In other words $\nabla \mathbf{A}$ is how fast $\mathbf{A}$ changes as you move through space.

in rectangular coordinates:
$$\nabla \mathbf{A} = \hat{x}\frac{\partial A}{\partial x} + \hat{y}\frac{\partial A}{\partial y} + \hat{z}\frac{\partial A}{\partial z}$$

in cylindrical coordinates:
$$\nabla \mathbf{A} = \hat{r}\frac{\partial A}{\partial r} + \hat{\phi}\frac{1}{r}\frac{\partial A}{\partial \phi} + \hat{z}\frac{\partial A}{\partial z}$$

in spherical coordinates:
$$\nabla \mathbf{A} = \hat{r}\frac{\partial A}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial A}{\partial \theta} + \hat{\phi}\frac{1}{r\sin\theta}\frac{\partial A}{\partial \phi}$$

∇ GRADIENT

$$\boxed{\nabla\vec{\Phi} = -\mathbf{E}}$$

"The gradient of the vector Φ" or "del Φ" is equal to the negative of the electric field vector.

$\nabla\Phi$ is a vector giving the direction and magnitude of the maximum spatial variation of the scalar function Φ at a point in space.

$$\nabla\vec{\Phi} = \hat{\mathbf{x}}\frac{\partial \Phi}{\partial x} + \hat{\mathbf{y}}\frac{\partial \Phi}{\partial y} + \hat{\mathbf{z}}\frac{\partial \Phi}{\partial z}$$

Ñ× DIVERGENCE

$\nabla\cdot$ is also a vector operator, combining the "del" or "grad" operator with the dot product operator and is read as "the divergence of". In this form of Gauss' law, where $\mathbf{D}$ is a density per unit area, with the operators applied, $\nabla\cdot\mathbf{D}$ becomes a density per unit volume.

$$\boxed{\text{div } \mathbf{D} = \nabla \cdot \mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho}$$

$\mathbf{D}$ = electric flux density vector $\mathbf{D} = \varepsilon\mathbf{E}$ [C/m^2]
ρ = source charge density [C/m^3]

$\tilde{N}^2$ THE LAPLACIAN

∇^2 is a combination of the divergence and del operations, i.e. $\text{div}(\text{grad } \Phi) = \nabla \cdot \nabla \, \Phi = \nabla^2 \, \Phi$. It is read as "the LaPlacian of" or "del squared".

$$\nabla^2 \mathbf{F} = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2}$$

$\mathbf{F}$ = electric potential [V]

GRAPHING TERMINOLOGY

With *x* being the horizontal axis and *y* the vertical, we have a graph of *y* **versus** *x* or *y* **as a function of** *x*. The *x*-axis represents the **independent variable** and the *y*-axis represents the **dependent variable**, so that when a graph is used to illustrate data, the data of regular interval (often this is time) is plotted on the *x*-axis and the corresponding data is dependent on those values and is plotted on the *y*-axis.

HYPERBOLIC FUNCTIONS

$$j \sin \theta = \sinh(j\theta)$$

$$j \cos \theta = \cosh(j\theta)$$

$$j \tan \theta = \tanh(j\theta)$$

TAYLOR SERIES

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x, \quad x \ll 1$$

$$\frac{1}{1-x^2} \approx 1 + x^2 + x^4 + x^6 + \cdots, \quad |x| < 1$$

$$\frac{1}{1 \pm x} \approx 1 \mp x, \quad x \ll 1$$

ELECTROMAGNETIC SPECTRUM

FREQUENCY	WAVELENGTH (free space)	DESIGNATION	APPLICATIONS
< 3 Hz	> 100 Mm		Geophysical prospecting
3-30 Hz	10-100 Mm	ELF	Detection of buried metals
30-300 Hz	1-10 Mm	SLF	Power transmission, submarine communications
0.3-3 kHz	0.1-1 Mm	ULF	Telephone, audio
3-30 kHz	10-100 km	VLF	Navigation, positioning, naval communications
30-300 kHz	1-10 km	LF	Navigation, radio beacons
0.3-3 MHz	0.1-1 km	MF	AM broadcasting
3-30 MHz	10-100 m	HF	Short wave, citizens' band
30-300 MHz 54-72 76-88 88-108 174-216	1-10 m	VHF	TV, FM, police TV channels 2-4 TV channels 5-6 FM radio TV channels 7-13
0.3-3 GHz 470-890 MHz 915 MHz 800-2500 MHz 1-2 2.45 2-4	10-100 cm	UHF "money band"	Radar, TV, GPS, cellular phone TV channels 14-83 Microwave ovens (Europe) PCS cellular phones, analog at 900 MHz, GSM/CDMA at 1900 L-band, GPS system Microwave ovens (U.S.) S-band
3-30 GHz 4-8 8-12 12-18 18-27	1-10 cm	SHF	Radar, satellite communications C-band X-band (Police radar at 11 GHz) K_u-band (dBS Primestar at 14 GHz) K-band (Police radar at 22 GHz)
30-300 GHz 27-40 40-60 60-80 80-100	0.1-1 cm	EHF	Radar, remote sensing K_a-band (Police radar at 35 GHz) U-band V-band W-band
0.3-1 THz	0.3-1 mm	Millimeter	Astromony, meteorology
10^{12}-10^{14} Hz	3-300 μm	Infrared	Heating, night vision, optical communications
3.95×10^{14}-7.7×10^{14} Hz	390-760 nm 625-760 600-625 577-600 492-577 455-492 390-455	Visible light	Vision, astronomy, optical communications Red Orange Yellow Green Blue Violet
10^{15}-10^{18} Hz	0.3-300 nm	Ultraviolet	Sterilization
10^{16}-10^{21} Hz		X-rays	Medical diagnosis
10^{18}-10^{22} Hz		γ-rays	Cancer therapy, astrophysics
> 10^{22} Hz		Cosmic rays	Astrophysics

GLOSSARY

anisotropic materials materials in which the electric polarization vector is not in the same direction as the electric field. The values of ε, μ, and σ are dependent on the field direction. Examples are crystal structures and ionized gases.

complex permittivity e The imaginary part accounts for heat loss in the medium due to damping of the vibrating dipole moments.

dielectric An insulator. When the presence of an applied field displaces electrons within a molecule away from their average positions, the material is said to be polarized. When we consider the polarizations of insulators, we refer to them as *dielectrics*.

empirical A result based on observation or experience rather than theory, e.g. *empirical data*, *empirical formulas*. Capable of being verified or disproved by observation or experiment, e.g. *empirical laws*.

evanescent wave A wave for which β=0. α will be negative. That is, γ is purely real. The wave has infinite wavelength—there is no oscillation.

isotropic materials materials in which the electric polarization vector is in the same direction as the electric field. The material responds in the same way for all directions of an electric field vector, i.e. the values of ε, μ, and σ are constant regardless of the field direction.

linear materials materials which respond proportionally to increased field levels. The value of μ is not related to **H** and the value of ε is not related to **E**. Glass is linear, iron is non-linear.

overdamped system in the case of a transmission line, this means that when the source voltage is applied the line voltage rises to the final voltage without exceeding it.

time variable materials materials whose response to an electric field changes over time, e.g. when a sound wave passes through them.

transverse plane perpendicular, e.g. the *x-y* plane is *transverse* to *z*.

underdamped system in the case of a transmission line, this means that after the source voltage is applied the line voltage periodically exceeds the final voltage.

wave number *k* The phase constant for the uniform plane wave. *k* may be considered a constant of the medium at a particular frequency.

Ixia Special Edition

5G

for dummies®

A Wiley Brand

- Innovations that make 5G possible
- Achieve scale through network virtualization
- See the future of the IoT and New Radio

Compliments of

ixia

Kalyan Sundhar

Lawrence C. Miller

About Ixia

Ixia, recently acquired by Keysight Technologies, provides testing, visibility, and security solutions, strengthening applications across networks and cloud environments for enterprises, service providers, and network equipment manufacturers. Ixia offers companies trusted environments in which to develop, deploy, and operate. Customers worldwide rely on Ixia to verify their designs, optimize their performance, and ensure protection of their networks and cloud environments to make their applications stronger. Learn more at **www.ixiacom.com**.

About Keysight Technologies

Keysight Technologies is a leading technology company that helps its engineering, enterprise, and service provider customers optimize networks and bring electronic products to market faster and at a lower cost. Keysight's solutions go where the electronic signal goes, from design simulation, to prototype validation, to manufacturing test, to optimization in networks and cloud environments. Customers span the worldwide communications ecosystem, aerospace and defense, automotive, energy, semiconductor, and general electronics end markets. Keysight generated revenues of $2.9 billion in fiscal year 2016. In April 2017, Keysight acquired Ixia, a leader in network test, visibility, and security. More information is available at **www.keysight.com**.

5G

Ixia Special Edition

by Kalyan Sundhar and Lawrence C. Miller

for dummies®

A Wiley Brand

5G For Dummies®, Ixia Special Edition

Publisher's Acknowledgments

Table of Contents

Introduction

The next-generation mobile network (NGMN) is on the horizon. 5G, the next iteration of 4G Long Term Evolution (LTE) networks, will enable significantly greater mobile speeds — as much as 20 gigabits per second (Gbps) with less than one millisecond (ms) latency — to enable real-time connectivity for mission-critical and potentially lifesaving devices and applications. 5G will also provide truly ubiquitous connectivity in the most challenging and remote areas of the world whether on land, in the air, or at sea — even on the 42nd floor of an office building in downtown Chicago! Finally, 5G networks will connect billions of Internet of Things (IoT) devices with a wide variety of speed and data volume requirements.

But 5G is an ambitious goal. Work on key technologies to enable 5G has already begun. In much the same way that 4G LTE was rolled out in 2008, but is only now achieving the 4G LTE goal of 1 Gbps speeds with the 4G LTE Advanced standard, 5G will be a steady evolution that begins with commercial availability expected in 2020. Many technologies that have emerged in the evolution of 4G LTE, such as carrier aggregation (CA) and multiple input multiple output (MIMO), will continue to develop to achieve the massive speed and scale required in 5G. Innovative new technologies will leverage unlicensed spectrum — where Wi-Fi operates — to offload certain traffic from the carrier networks to create more capacity in their data pipes. Low-power technologies such as NarrowBand IoT (NB-IoT), LTE for Machines (LTE-M), Long Range Wide Area Network (LoRaWAN), Sigfox, and others, will be used in billions of IoT devices, and a new radio interface technology — 5G New Radio (5G NR) — will be developed for connections between User Equipment (UE) and carrier enhanced Node B (eNodeB) stations. Finally, carriers will fully embrace virtualization technologies in their core networks to enable massive scale and efficiency.

In this book, you learn about the technological innovations that are being developed today to enable a 5G future. You also learn about potential use cases that will transform entire businesses and industries, and create new business models and opportunities.

About This Book

5G For Dummies, Ixia Special Edition, consists of eight short chapters that explore

>> How wireless communications technology has evolved and where it's going next (Chapter 1).

>> How developments in today's networks are blazing the 5G trail to higher speeds (Chapter 2).

>> Which technologies in unlicensed spectrum will be leveraged in the 5G networks of the future (Chapter 3).

>> Why the Internet of Things requires 5G connectivity (Chapter 4).

>> Where virtualization in mobile networks can help address the need for scale and elasticity (Chapter 5).

>> What 5G New Radio (5G NR) is and how it will help create a 5G future (Chapter 6).

>> How 5G will be used in various use case scenarios (Chapter 7).

>> Ten common myths and the reality of 5G (Chapter 8).

Foolish Assumptions

It's been said that most assumptions have outlived their uselessness, but we assume a few things nonetheless!

Mainly, we assume that you either work in a technology profession or you're an avid user of wireless communications technology — if you have a smartphone within arm's distance of you, we're talking about you!

Beyond a basic knowledge of wireless communications and mobile technology in general, we don't assume you have a particularly strong technical background. As such, this book is written primarily for nontechnical readers — we explain any technical terms and concepts that come up in this book.

If any of these assumptions describe you, this book is for you! If none of these assumptions describe you, keep reading anyway. It's a great book, and when you finish reading it, you'll know enough about 5G to be dangerous!

Icons Used in This Book

Throughout this book, we occasionally use special icons to call attention to important information. Here's what to expect:

REMEMBER

This icon points out information you should commit to your non-volatile memory, your gray matter, or your noggin' – along with anniversaries and birthdays!

TECHNICAL STUFF

You won't find a map of the human genome here, but if you seek to attain the seventh level of NERD-vana, perk up! This icon explains the jargon beneath the jargon.

TIP

Tips are appreciated, never expected — and we sure hope you'll appreciate these tips! This icon points out useful nuggets of information.

Beyond the Book

There's only so much we can cover in 48 short pages, so if you find yourself at the end of this book thinking "gosh, this is a great book; where can I learn more?" just go to `www.ixiacom.com`.

Where to Go from Here

With our apologies to Lewis Carroll, Alice, and the Cheshire cat:

"Would you tell me, please, which way I ought to go from here?"

"That depends a good deal on where you want to get to," said the Cat — er, the Dummies Man.

"I don't much care where . . . ," said Alice.

"Then it doesn't matter which way you go!"

That's certainly true of *5G For Dummies*, which, like Alice in Wonderland, is also destined to become a timeless classic!

If you don't know where you're going, any chapter will get you there — but Chapter 1 might be a good place to start! However, if you see a particular topic that piques your interest, feel free to jump ahead to that chapter. Each chapter is written to stand on its own, so you can read this book in any order that suits you (though we don't recommend upside down or backward).

We promise you won't get lost falling down the rabbit hole!

Chapter **1**

Understanding the Journey to a 5G Future

n this chapter, you take a glimpse back at the evolution of wireless communications and a look ahead to the 5G future.

Tracing the Evolution of Wireless Communications

For more than a century, radio technology has been enabling wireless communications over ever greater distances and with ever greater capabilities.

In the late nineteenth century, Guglielmo Marconi built the first wireless telegraphy system, capable of transmitting Morse code via radio signals up to one-half mile. Today, more than seven billion mobile devices enable us to communicate with anyone, anywhere in the world.

The first truly mobile two-way radio was developed in 1923 and used in Australian police cars — although it took up the entire back seat of a patrol car. Hand-held radios — "walkie-talkies" — were first used in World War II.

In 1973, the first call on a hand-held cellular phone was made — the cellular phone was described as a "brick" weighing nearly two pounds, with just 30 minutes of talk time and a ten-hour battery recharge time. Ten years later, Motorola introduced the DynaTAC phone, weighing just one pound and costing $3,500.

To support modern wireless communications, cellular networks have evolved over several generations, as follows:

>> **1G (analog cellular):** The first analog cellular service was launched in Japan in 1979. In 1983, the Advanced Mobile Phone Service (AMPS) was launched in North America. Analog cellular signals permitted only voice traffic and were not encrypted, so they could be easily intercepted.

TECHNICAL STUFF

1G service consumed lots of spectrum and used the frequency division multiple access (FDMA) channel access method. FDMA allocates one or more frequency bands (or channels) to a user for communication.

>> **2G (digital cellular):** The second generation of cellular technology was launched in 1991 with the commercial release of the Global Standard for Mobile Communications (GSM) in Finland. Major innovations in 2G networks included:

- **Digital:** Digital signals generally have less static and background noise, and they use available spectrum more efficiently than do analog signals.

- **Encryption:** 2G digital calls can be encrypted to make eavesdropping and intercept more difficult.

- **Data:** Short message service (SMS) text messages were first introduced in 2G networks — O-M-2G!

TECHNICAL STUFF

2G technologies use either time division multiple access (TDMA) or code division multiple access (CDMA) channel access methods. TDMA divides a signal into different time slots, enabling multiple callers to share the same frequency channel. CDMA assigns a code to each caller and uses spread-spectrum technology to create a signal with a wider bandwidth.

In 2000, the European Telecommunications Standards Institute (ETSI) created the General Packet Radio Service (GPRS), which implemented packet-switched domains, in addition to existing circuit-switched domains. GPRS was

dubbed "2.5G" and had nothing to do with *Two and a Half Men,* introduced by CBS three years later.

» **3G (data driven):** Apple and Google brought smartphones to the masses with iPhones and Android devices, respectively, in the early 21st century. These powerful devices — and the mobile apps installed on them (including Global Positioning System or GPS, location-based services, and on-demand video) — created an insatiable appetite for faster download speeds. The first 3G networks, introduced in 1998, provided minimum information transfer rates of 200 kilobits per second (Kbps). The International Telecommunication Union (ITU) has never formally defined a standard for 3G data rates, so downlink data speeds vary widely — from 384 Kbps in a moving vehicle for Wideband Code Division Multiple Access (W-CDMA) to 42.2 megabits per second (Mbps) for Evolved High Speed Packet Access (HSPA+), also known as 3.5G, and 168 Mbps for Advanced HSPA+.

» **4G (Long Term Evolution):** Commercially available 4G mobile networks were rolled out in 2008, and 4G LTE followed in 2010. However, unlike 3G, the ITU Radiocommunication Sector (ITU-R) defined minimum 4G standards — but neither "4G" nor "4G LTE" meets those standards! The ITU-R International Mobile Telecommunications Advanced (IMT-Advanced) requirements include (among other things):

- Packet-switched all-IP core networks

- Peak data rates of approximately 100 Mbps for high mobility (such as moving vehicles)

- Peak data rates of approximately 1 gigabit per second (Gbps) for low mobility (such as walking — or your authors sprinting)

With the introduction of LTE Advanced, true 4G speeds of up to 1 Gbps finally arrived. LTE Advanced Pro is the next evolution of LTE technology, and it establishes the foundation for 5G. LTE Advanced Pro will deliver speeds in excess of 3 Gbps with less than 2 milliseconds (ms) of latency.

REMEMBER

At this point you may be thinking, "Long Term Evolution — no kidding! Will 5G ever get here?" However, the trend has been for each generation of mobile technology innovation to take about a decade — most of us just didn't pay attention before we got our first smartphones midway through the 3G era.

Focusing on the 5G Vision

The vision for the 5G future is bold: It is much more than just the next iteration of mobile networks. 5G will achieve three main goals:

» **Speed (ultra-high speed radio access):** 5G will provide download speeds of up to 20 Gbps. If you're wondering "Why would anyone ever need that much speed?" first answer this question: When have you ever heard anyone complain that their phone was *too fast*? It's also important to remember that bandwidth is shared by all the users on a cell tower. Today, if a few users are streaming a video at the airport or watching replays of a touchdown in a stadium, chances are that the download is choppy with lots of buffering, and the experience isn't so great. With 5G, you could theoretically download a 40 gigabyte (GB) 4K Ultra-High-Definition (UHD) movie (like *Jaws*) in less than a minute — you're gonna need a bigger data plan!

» **Responsiveness (ultra-low latency):** 5G networks will be used to control autonomous cars and high precision, mission-critical industrial devices in real-time. High reliability and availability at all times is a necessity for these use cases. For this to happen safely, end-to-end latency — the time it takes for data or commands to travel across the network — has to be extremely low. Latency in 5G networks will be five times faster than today's networks — less than 1ms.

» **Scale (massive connectivity):** By 2020, Gartner conservatively forecasts that there will be more than 21 billion connected devices in the Internet of Things (IoT). Some estimates predict more than 50 billion connected IoT devices by 2020. That's anywhere from three to seven connected devices for every person on the planet in 2020 — not including smartphones, tablets, and computers! These devices will have widely varying network requirements — from environmental sensors for agricultural applications installed in remote areas that might send a few bits of data every few days or weeks, to extremely high-precision, low-latency devices in nanobiotechnology, autonomous cars, and mission-critical industrial environments that rely on real-time communication for potentially lifesaving functions. 5G networks will need to handle the massive scale and

challenging connectivity requirements for all these diverse devices, sensors, and applications.

Ordering Up 5G in Five Easy Pieces

5G will be a quantum leap beyond today's networks and will require many technological innovations. We serve them up here in five easy pieces (easier than making wheat toast for Jack Nich-olson — "hold the butter, the lettuce, the mayonnaise . . . and the chicken") in this section and the chapters that follow (see Figure 1-1):

>> **Speeds and feeds.** Speeds of up to 20 Gbps will be achieved using a combination of innovations such as carrier aggrega-tion (CA), massive multiple input multiple output (MIMO), and quadrature amplitude modulation (QAM). You learn about speeds and feeds in Chapter 2.

>> **Unlicensed spectrum:** MNOs are increasingly using unlicensed spectrum in the 2.4 and 5 Gigahertz (GHz) frequency bands. 5G networks will need to tap into the vast amount of spectrum available in these unlicensed bands to offload traffic in heavily congested areas and provide connectivity for billions of IoT devices. Advancements in Wi-Fi, LTE in Unlicensed spectrum (LTE-U), License Assisted Access (LAA), and MulteFire, among others, provide better quality and regulated access to unlicensed spectrum. You learn about these advancements in Chapter 3.

>> **Internet of Things (IoT):** IoT devices pose a diverse set of requirements and challenges for 5G networks. It's only fair that IoT should likewise pose a diverse set of solutions as well! You learn about a few of these solutions — including NarrowBand IoT (NB-IoT), LTE Category M1 (LTE-M), Long Range (LoRa) and Sigfox — in Chapter 4.

>> **Virtualization:** Network functions virtualization (NFV) enables the massive scale and rapid elasticity that MNOs will require in their 5G networks. Virtualization enables a virtual evolved packet core (vEPC), centralized radio access network (C-RAN), mobile edge computing (MEC), and network slicing — all explained in Chapter 5.

» **New Radio (NR):** Although the other 5G innovations introduced in this section all have strong starting points in LTE Advanced Pro, 5G NR is a true 5G native technology that has yet to be standardized. 5G NR addresses the need for a new radio access technology that will enable access speeds up to 20 Gbps — and you learn about it in Chapter 6!
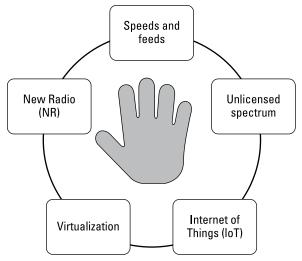
FIGURE 1-1: The five keys to 5G.

Chapter **2**

Achieving Faster Speeds and Larger Feeds

Over the past decade, the need for speed in mobile networks has increased dramatically. To address this need, 5G will increase the speeds of today's most advanced Long Term Evolution (LTE) networks by an order of magnitude — from a few gigabits per second (Gbps) to as much as 20 Gbps. In this chapter, you get a glimpse of the engine — or, more correctly, the parts and components of the engine — that will power the 5G networks of the future.

Fattening the Data Pipe

Wireless spectrum is limited and highly regulated throughout the world. The International Telecommunication Union (ITU) allocates frequency spectrum worldwide, and governing bodies within the respective countries then license that spectrum for use by individual mobile network operators (MNOs). The ITU has identified three International Telecommunication Regions

throughout the world with distinct frequency bands allocated to each region (see Figure 2-1).
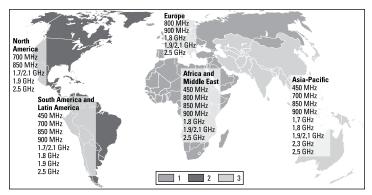
FIGURE 2-1: ITU International Telecommunication Regions and spectrum allocations.

MNOs must pay for the right to transmit and receive data using this shared medium in each country in which they operate. To fatten a wireless data pipe, an MNO must procure additional licenses to use a designated frequency spectrum.

In the U.S., the Federal Communications Commission (FCC) auctions spectrum to MNOs — and we aren't talking about eBay auctions here! In the U.S. alone, spectrum auctions have raised more than 60 billion dollars since 1994. Another challenge for MNOs is that available spectrum is limited and going once . . . going twice . . . sold!

Thus, MNOs must find new ways to use existing spectrum more efficiently. LTE Advanced and LTE Advanced Pro are pioneering the path to 5G, using several innovations to get more data through existing spectrum, including:

» Carrier aggregation (CA)

» Multiple input multiple output (MIMO)

» Quadrature amplitude modulation (QAM)

We All Bundle — with CA

As you might imagine, the process of buying and selling spectrum, over time, causes spectrum to be sliced and diced in some pretty creative ways. MNOs also come and go, or get merged, acquired, and divested (for example, Cingular Wireless and MCI). All of this causes yet another challenge — contiguous spectrum is hard to find and MNOs must cobble together different bands to maximize their available bandwidth. It's sort of like a business being so unreasonable as to want a continuous range of direct inward dialing (DID) phone numbers to simplify its phone switch programming, company directories, and business cards! Except that in the case of frequency spectrum, non-contiguous bands aren't just a messy inconvenience — they limit available bandwidth.

For LTE networks, including LTE Advanced and LTE Advanced Pro, four carrier bandwidths (or sizes) are available for transporting data:

» 5 megahertz (MHz)

» 10 MHz

» 15 MHz

» 20 MHz

Larger bandwidths can transport more data. For example, 10 MHz can transport data at 37.5 megabits per second (Mbps) and 20 MHz can transport data at 75 Mbps. These data transfer rates assume a single antenna on the user equipment (UE) side and on the Evolved Node B (eNodeB) side. This is known as single input single output (SISO).

TECHNICAL STUFF

User equipment (UE) refers to an end-user device in a mobile network, such as a smartphone. *Evolved Node B (eNodeB)* is the MNO hardware — for example, a base transceiver station (BTS) — that wirelessly communicates directly with the UE.

Carrier aggregation (CA) is a technique that allows an MNO to use more than one component carrier (CC) — known as the *secondary carrier* — as an additional data pipe. For example, 2CA enables any of the four carrier bandwidths (5, 10, 15, or 20 MHz) to be used as the primary carrier and any of the other four carrier bandwidths to be used as the secondary carrier.

In the simplest deployment scenario, known as *intra-band contiguous CA*, a 20 MHz primary and a 20 MHz secondary carrier (totaling 40 MHz) would provide twice the maximum possible bandwidth and throughput of a single carrier.

For SISO, a 2CA of 20+20 MHz would provide 150 Mbps of data throughput.

TIP MNOs in the U.S. are currently deploying 3CA and some are already moving to 4CA (see Figure 2-2). The challenge for MNOs now is to find three or more CCs that they own the license to operate in and can aggregate. The LTE Advanced standard specifies 5CA totaling 100 MHz, while LTE Advanced Pro calls for 32CA totaling 640 MHz of aggregated carrier bandwidth.
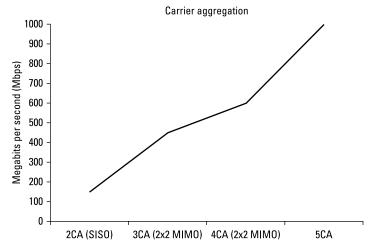
FIGURE 2-2: Carrier aggregation (CA) increases available bandwidth and data throughput.

Eeny, Meeny, Miny, MIMO

One of the techniques for fattening the data pipe defined from day one in 4G LTE networks is multiple input multiple output (MIMO) — using multiple antennas on the transmit and receive side in the wireless network. Using a technique called *spatial multiplexing*, it's possible to send a different data stream on each antenna, thereby increasing the throughput to the cell tower and to the user device.

While 2x2 MIMO deployments (two antennas in, two antennas out) are common today, 4x4 MIMO (four antennas in, four antennas out) is becoming more practical. 4x4 MIMO effectively enables four independent data streams, thereby increasing the throughput by four times. Thus, the 2CA deployment example (described in the previous section) with two 20 MHz channels and a 2x2 MIMO configuration would increase the data throughput from 150 Mbps to 300 Mbps. The same scenario with a 4x4 MIMO configuration would theoretically increase data throughput to 600 Mbps.

LTE Advanced specifications for MIMO accommodate 8-, 16-, and 32-antenna configurations. LTE Advanced Pro specifications increase MIMO configurations to 64 antenna elements. Realistically, using 64 antenna streams to increase the throughput of a single UE isn't practical. However, using a technique called *beamforming,* the eNodeB can focus energy (that is, steer beams of data) to a particular UE, thereby increasing the throughput of that UE while simultaneously handling other UEs through a different set of beams. Massive MIMO will utilize massive antenna arrays — comprised of hundreds of antennas — to provide efficient signal coverage and higher data rates with lower latency in 5G networks.

No Qualms About QAM

Quadrature amplitude modulation (QAM) is a technique widely used to vary data signals on a carrier used for radio communication. When used for digital transmission of radio communication applications, QAM can carry higher data rates than ordinary amplitude modulated and phase modulated schemes. In QAM, the constellation points are normally arranged in a square grid with equal vertical and horizontal spacing. As a result, the most common forms of QAM use a constellation with the number of points equal to a power of 2 (such as 4, 16, and 64). Thus, 16-QAM uses a 16-point constellation while 256-QAM uses a 256-point constellation.

By using higher order modulation formats (that is, more points on the constellation), it is possible to transmit more bits per symbol. So, 64-QAM has six bits per symbol (more data transmitted), whereas 16-QAM uses only four bits per symbol (less data transmitted). However, the points for a higher QAM are closer together and are therefore more susceptible to noise and data errors. 256-QAM has been used for data in digital cable

communications and is now starting to be used for radio communications. 256-QAM is included as part of the LTE Advanced Release 12 standard from the Third Generation Partnership Project (3GPP) because it is likely to work with small cell towers. T-Mobile has achieved 400Mbps downlink speeds in trials through a combination of 4x4 MIMO and 256-QAM.

REMEMBER

5CA and beyond, massive MIMO, and higher-order QAM are all techniques defined as part of the LTE Advanced and LTE Advance Pro standards. With LTE Advanced deployments in full swing, 1 Gbps download speeds are getting closer to becoming a reality in operational networks. 5G will use all these techniques to achieve 20 Gbps data speeds. For example, these are a few of the developments on the 5G horizon:

» CA on very large bandwidths, leading to aggregate totals of 800 MHz (and even 1 GHz) carrier bandwidth

» Massive MIMO antenna arrays of 128 elements or more

» 1024-QAM, and even 4096-QAM

Chapter **3**

Tapping into Unlicensed Spectrum

U nlicensed spectrum is used by low-power devices to transmit and receive wireless signals over short distances — typically a few meters. Although specific devices are permitted to operate only in specific bands, the process of getting certifications is not time-consuming or costly, compared to getting a cellular operator license. Some common devices in this category include garage door openers, nursery monitors, home security systems, cordless phones, and Bluetooth speakers/headsets. In this chapter, you learn about the role of unlicensed spectrum in the 5G future.

Giving a "High 5(G)" to Wi-Fi Advancements

Wi-Fi operates in the unlicensed 2.4 gigahertz (GHz) and 5 GHz spectrums. There are fewer rules on who can access these bands and more available spectrum compared to licensed frequency bands. Wi-Fi devices must therefore compete to use the same spectrum as other devices.

Over the past few years, improvements in speed and the advancement of new capabilities have made Wi-Fi viable (and lucrative) for many mobile network operators (MNOs).

MNOs now regularly use Wi-Fi to offload their cellular networks wherever possible. Newer Wi-Fi standards enable massive speed increases. The standards utilize many of the latest multiple input multiple output (MIMO) beamforming techniques and higher quadrature amplitude modulation (QAM) schemes as those used in the licensed spectrum (discussed in Chapter 2).

For example, 802.11ac delivers Wi-Fi speeds of up to several gigabits per second (Gbps) operating in the 5 GHz band, using 80 or 160 megahertz (MHz) wide channels, advanced beamforming techniques, eight spatial streams (MIMO), multi-user MIMO (MU-MIMO), and 256-QAM (which produces four times the spectral efficiency of the previous 802.11n Wi-Fi standard).

The next wave in Wi-Fi is 802.11ax. 802.11ax will use orthogonal frequency-division multiple access (OFDMA) — the same technique used in Long Term Evolution (LTE) networks — in which different subcarriers within a carrier can be used to transport data for different users. As a result, more than one user equipment (UE) device can access the same medium at a given instant without having to back off or concede the medium to another UE device.

TIP In case you're wondering, other standards exist between 802.11ac and 802.11ax, but they define Wi-Fi standards in unlicensed spectrum other than 2.4 GHz and 5 GHz — and they don't sound as cool as "A-X"!

LTE in Unlicensed Spectrum

LTE in Unlicensed spectrum (LTE-U) is a proposal that was originally developed by Qualcomm for LTE to co-exist with Wi-Fi in shared unlicensed spectrum. In LTE-U, calls are initially set up using licensed LTE spectrum. Additional carriers (for data) can then be aggregated from the unlicensed spectrum. This method will allow the operators to use unlicensed spectrum to "fatten the data pipe."

The Wi-Fi Alliance developed a Wi-Fi co-existence test plan to ensure that Wi-Fi and LTE could peacefully co-exist, and MNOs (including T-Mobile, AT&T, and Verizon) began experimenting with unlicensed spectrum in addition to their licensed spectrum.

LTE-U has also generated a lot of interest from a small cell perspective because this approach potentially makes small cells a viable alternative to Wi-Fi hot spots.

As you might imagine, strong reservations about LTE-U exist, because of concerns that it will unfairly use the unlicensed spectrum and potentially interfere with other Wi-Fi users. Cable companies, such as Comcast, Charter Communications, and Time Warner Cable (TWC), as well as Google and Microsoft, are opposed to LTE-U. As of February 2017, the U.S. Federal Communications Commission (FCC) has approved Ericsson and Nokia equipment as LTE-U certified.

License Assisted Access

Unlike LTE-U, License Assisted Access (LAA) is a standard defined in the Release 13 specification from the Third Generation Partnership Project (3GPP). LAA attempts to resolve contention issues with Wi-Fi in unlicensed spectrum using a protocol many of us learned as children — Listen Before Talk (LBT). The LBT contention protocol requires LAA to listen to the unlicensed channel first, and then, if there is no other active Wi-Fi user (the channel is silent), the LTE user can use the channel.

TIP

A recent report by Strategy Analytics states that only 16 percent of operators can achieve gigabit LTE without using unlicensed spectrum. Using LTE-U and LAA, 64 percent of operators can achieve gigabit LTE.

Qualcomm's over-the-air trials with Deutsche Telecom in Germany showed that LAA had increased coverage and capacity compared to Wi-Fi on the same spectrum. By respecting the LBT maxim (uh, protocol), the behavior of LAA is more closely modeled to the principle "love thy neighbor" and is a better neighbor to Wi-Fi than Wi-Fi is to LTE. Wi-Fi was more of an imp during its formative years and ignores LBT, trying to access the same medium at the same time as others.

Enhanced LAA (eLAA), which aggregates the uplink on unlicensed spectrum, is the next step in the evolution of LAA. eLAA will be defined in the 3GPP Release 14 specification.

REMEMBER

Both LAA and eLAA require the initial call to be set up on LTE licensed spectrum as the primary channel. Unlicensed spectrum is only used to "fatten" the data pipes.

MulteFire

MulteFire goes beyond LTE-U and LAA by enabling the LTE primary channel on unlicensed spectrum. In fact, MulteFire uses unlicensed spectrum exclusively. This means MulteFire can be deployed for LTE by anyone — without owning licensed spectrum — such as Internet service providers (ISPs) and commercial enterprises. MulteFire also benefits MNOs by providing new deployment opportunities for offloading and augmenting their cellular networks.

In January 2017, the MulteFire Alliance released version 1.0, defining an LTE-like network that can run entirely on unlicensed spectrum and, in some cases, may be an alternative to Wi-Fi with more capacity, better security, and easier handoffs across carrier networks.

REMEMBER

Today's 4G spectrum has around 500 MHz of unlicensed spectrum available to deploy LTE-U, LAA, and MulteFire. In July 2016, the FCC opened up 5G spectrum in the U.S., which has 10.85 GHz of total spectrum (3.85 GHz of licensed spectrum and 7 GHz of unlicensed spectrum). The techniques that are now evolving in LTE Advanced Pro (Releases 13 and 14) — LTE-U, LAA (and eLAA), MulteFire, and others (see Figure 3-1) — will be vital as even more unlicensed spectrum becomes available in 5G. The key goal will be to provide users with a seamless experience, irrespective of whether they are operating on a licensed or unlicensed band.
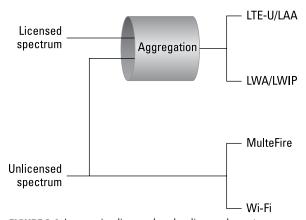
FIGURE 3-1: Leveraging licensed and unlicensed spectrum.

Chapter **4**
Enabling Massive IoT

When 4G networks were introduced in 2008, there were close to 700 million mobile subscribers worldwide. Today, there are more than 7 billion mobile subscribers worldwide. By 2020, according to the most conservative estimates, there will be approximately 20 billion Internet of Things (IoT) devices, in addition to the 9 billion mobile subscribers that are expected worldwide.

Thus, 5G networks must not only deliver extreme speeds and feeds, they must provide massive scale, predictability, and reliability to eventually support as many as a trillion connected devices, including mission-critical and potentially lifesaving applications and scenarios. In this chapter, you learn about the key requirements for IoT device connectivity and the innovations that are being developed to support IoT.

Key Connectivity Requirements for IoT Devices

Unlike smartphones and other cellular devices, IoT device communications can be sporadic in nature. Many of these devices "sleep" (to extend battery life for ten or more years in some cases)

for long periods of time — hours, days, or weeks — before transmitting a few bytes of data, and thus needn't always be connected to the network. 5G networks must be designed to handle infrequent, but important communications from these types of IoT devices. Although the amount of data these devices send may be significantly lower, they may still be of a time-critical nature. For example, a sensor that detects a hazardous condition may instruct an Evolved Node B (eNodeB) element to shut down equipment in an industrial plant or building. These types of communication, though extremely rare, must be handled with the utmost responsiveness and reliability.

Additionally, eNodeB elements will require massive capacity to scale to support tens of thousands of IoT devices — all with different communications requirements and characteristics — in a single cell.

Finally, a wide variety of security threats and new attack vectors will be surfaced in IoT devices. Unlike many of today's threats — identity theft and credit card fraud, for example — many IoT security threats will be potentially life threatening. Future threats may include hacked control systems in autonomous vehicles and smart grids, or compromised medical devices such as insulin pumps and pacemakers. Thus, 5G technologies will need to provide secure end-to-end communications throughout the network.

NarrowBand IoT (NB-IoT)

The Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) Advanced Pro Release 13 specification defines a new standard — NarrowBand IoT (NB-IoT) — for handling low volumes of data (similar to 2G) from tens of thousands of devices in a single cell (tower, not amoeba).

NB-IoT extends LTE to even narrower bandwidths optimized for low data rate, latency-tolerant IoT applications. NB-IoT reduces device complexity, enables multiyear battery life, and provides deeper coverage to reach sensors in challenging locations, such as remote rural areas or inside buildings.

Perhaps one of the most attractive features of NB-IoT is its ability to leverage already ubiquitous LTE networks, in addition to providing high quality of service (QoS) and comprehensive security.

NB-IoT can be deployed in three different modes:

- **»** **Standalone as a dedicated carrier:** Can use GSM frequencies in a bandwidth of 200 kHz. This does not require LTE.
- **»** **Guard band:** Uses a free resource block within the LTE guard band. This allows the IoT devices to not compete with other LTE devices for the resource blocks within the carrier.
- **»** **LTE in-band:** Uses a resource block within the LTE frequency band. The rest of the blocks are used by the regular LTE devices.

Long Term Evolution for Machines (LTE-M)

Like NB-IoT, Long Term Evolution for Machines (LTE-M or LTE Category M1) leverages a narrow slice of existing LTE spectrum to send and receive data for IoT devices. LTE-M has the same benefits as NB-IoT, but uses a larger network slice than NB-IoT (1.4 MHz compared to 180 KHz in NB-IoT) and leverages the LTE protocol more than the NB-IoT in terms of reusing the same control, data, and transport channels.

TIP Verizon launched the first LTE-M network in the U.S. on March 31, 2017, and ATT Wireless was expected to follow shortly after.

LoRaWAN and Sigfox

Long Range Wide Area Network (LoRaWAN) is a Low Power Wide Area Network (LPWAN) specification for wireless, battery-operated IoT devices. LoRaWAN operates in the sub-1 GHz unlicensed spectrum band. This limits the volume and frequency of traffic, as well as the ability of the base station to control the network and send traffic down. However, LoRaWAN has great advantages in terms of battery life and cost, and communication is bi-directional.

Sigfox is a French company that created a technology similar to LoRaWAN for IoT device communication. Sigfox technology uses

very low bandwidth connections. It is not bi-directional and is only used for sending sparse uplink data with very limited down-link. Like LoRaWAN, Sigfox operates in the sub-1 GHz space and thus uses very low power.

LoRaWAN and Sigfox will be used with certain types of sensors, smart meters, and other low data IoT devices. A disadvantage for these technologies is that, unlike NB-IoT, which is built on top of existing LTE infrastructure, LoRaWAN and Sigfox are not inte-grated with LTE. However, the cost of deploying a LoRaWAN or Sigfox based IoT device is far less than for an NB-IoT or LTE-M device — by an order of magnitude, since NB-IoT and LTE-M devices must integrate LTE modules into their devices.

REMEMBER

As countries and mobile network operators (MNOs) make IoT technology decisions, it is likely that many standards, protocols, and technologies will need to co-exist. Factors such as cost, cov-erage, battery life, and integration with older 4G networks will determine which standards, protocols, and technologies are best for each use case.

The 3GPP specifications for IoT technologies are just starting to come out as part of the LTE-A Pro standards. Many IoT devices will need to operate at very low power, ideally suited for sub-1 GHz spectrum rather than the millimeter wave (mmWave) spectrum (discussed in Chapter 6). That said, there are some compelling 5G techniques, like resource spread multiple access (RSMA) wave-forms that allow grant-free transmissions for "things" to send their data without prior scheduling by the eNodeB. As a result, the scheduling algorithm becomes less complex. 5G will also enable multi-hop mesh for these low-power devices, allowing out-of-coverage devices to relay to other connected devices in order to send data to the eNodeB.

Chapter **5**

Getting Real About the Need to Virtualize

Virtualization has been a hot topic in the technology industry for many years and its advantages transcend the cellular industry. In this chapter, you learn about the network elements that can be virtualized and the essential role of virtualization in 5G.

Driving 5G and IoT with Virtualization

Telecommunications companies and mobile network operators (MNOs) have invested heavily in their 4G and Long Term Evolution (LTE) networks. These organizations have embraced virtualization to enable faster, more agile, and scalable deployments that can keep pace with the explosion in subscriber data traffic and consumption, all while keeping control of their overall capital and operating costs. Even though these investments have already given users mobile connectivity of unprecedented speed and pervasiveness, they have only laid the foundation of what's to come, as 5G and smart, connected devices start to roll out.

A recent study by a division of Nokia Bell Labs provides a glimpse of what's coming. The study found that the number of Internet of Things (IoT) connected devices is expected to expand from 1.6 billion in 2014 to well over 20 billion by 2020. Also, by 2020, global consumption demand for digital content and services on portable devices will see an average increase of 30 to 45 times the

levels seen in 2014. Thus, MNOs will need to further accelerate their technology investments to meet ever-increasing consumer and business connectivity demands.

Recognizing the IoT explosion

Although MNOs in the U.S. are starting to embrace "pre-standard" 5G, vying for the first mover's advantage, the standards are not expected until 2018. This has echoes of the early days of 3G in the late 1990s and early 2000s, before proprietary implementations were standardized. It isn't surprising that some operators in the U.S. have already announced their plans to deploy 5G later this year, and some of the early adopters in Korea, Japan, and China are also ready to roll out 5G. Because of the rapid deployment, many 5G concepts are being solidified quickly.

5G is sometimes narrowly classified as a higher bandwidth radio access technology. But it is much more than that. It will also be the network for low-power devices and sensors that are classified as IoT devices, as well as low-latency applications. One example is the low latency required for some mission-critical devices, such as autonomous vehicles. The need for low latency will dictate that key LTE base station functions are distributed, with some moving closer to the edge and others being pooled in the cloud.

Another example is that vertical sectors will require different types of services from the 5G network. Some will need high band-width, while others will need low power. Some will need very low latency, and others will need very high availability. The vast volume of IoT devices will range from those that send multiple gigabits of data per second to those that will only send a few bits every month. This flexibility and elasticity can be supported only by advanced network virtualization.

Focusing on service

To support such varying 5G use cases across multiple verticals, MNOs need to shift from being network-centric to being more service-oriented.

To understand this shift, consider the concept of network slicing (see Figure 5-1). Here are a few typical cases:

» Mobile broadband with higher bandwidth video requiring high availability everywhere

>> Low-power sensors that can operate on a pair of AA batteries for 10 years

>> Autonomous vehicles that can zip through crowded city streets but brake within microseconds when they sense potential obstructions in their paths

TECHNICAL STUFF

Each of these scenarios requires a different configuration of the requirements and parameters in the network, potentially including:

>> Home Subscriber Server (HSS)

>> Mobility Management Entity (MME)

>> User equipment (UE)

>> Enhanced Node B (ENB or eNodeB)

>> Serving Gateway (SGW)

>> Packet Data Network Gateway (PGW)

>> Policy and Charging Rules Function (PCRF)

>> LTE Evolved Packet Core (EPC) interfaces (S1-MME, S11, S1-U, SGi)

In essence, each use case requires its own network slice. Networks must be built in a manner that allows speed, availability, capacity, and coverage to be allocated in logical slices to meet the demands of each use case.

The best way to implement network slices will be via virtualization — to use service provider software–defined networking (SDN), network functions virtualization (NFV), and network orchestration (see Figure 5-2). In each of these cases, the SDN controller will configure and build network slices that include service chains such as deep packet inspection (DPI), emails, and security scans per user or per service. The network services — including the Centralized Radio Access Network (C-RAN), Virtual Evolved Packet Core (vEPC), and Virtual IP Multimedia Subsystem (vIMS) — are virtualized as virtual network functions (VNFs) that allow MNOs to set up services rapidly, and scale them in response to network and service demands.
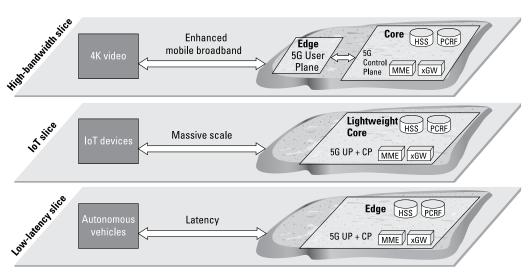
FIGURE 5-1: 5G network slices.
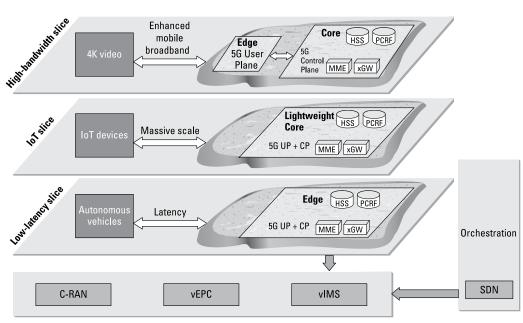
FIGURE 5-2: Network slices, VNFs, and management and orchestration in a 5G architecture.

Managing the migration

Thus, virtualization will be a vital piece of the puzzle as 5G is rolled out and the IoT grows. But the speed and agility that virtualization enables must be balanced against the need for network visibility, resilience, manageability and security, throughout the development, roll-out, and mass usage of each service.

MNOs and service providers must weigh the trade-offs between

>> Quality and cost.

>> Flexibility and control.

>> Moving too quickly and not moving fast enough.

REMEMBER As a result, the demand for simple, end-to-end solutions that can efficiently test and validate the effectiveness and benefits of virtualization at every stage is only going to increase. Network virtualization will be a key driver of the IoT ecosystem as it develops — and full-lifecycle reliability and performance testing will be a key element of this shift.

Virtualizing Network Components

When LTE networks were conceptualized, one notable difference from earlier 3G architectures was the idea of a central "brain" — the evolved Node B (eNodeB). From the eNodeB — the point at which the cellular network wirelessly connects directly to user equipment (UE) — to the Internet, the connectivity would all be Internet Protocol (IP) based.

When data centers and other wired network components (such as routers, switches, and firewalls) started being virtualized on standard off-the-shelf servers, it was only natural for the wireless, IP-based 4G evolved packet core (EPC) to follow.

Virtual EPC (vEPC)

Because packet gateways, policy servers, and subscriber databases are all IP-based in the LTE network, the EPC is an excellent

candidate for replacing proprietary hardware with virtualized functions hosted on standard, off-the-shelf servers. The benefit for MNOs is the elasticity this model enables — no more dedicated hardware and no need to oversubscribe every part of the network. Servers can be quickly and easily provisioned or de-provisioned based on real-time subscriber capacity needs. With SDN controllers and EPC functions migrated to NFV, and orchestration schemes added to manage it all, the wireless core could be fully virtualized and cloud ready!

Centralized Radio Access Network (C-RAN)

Virtualization in the radio access network (RAN) requires separating out latency-sensitive elements and the radio itself, then pooling the rest of the baseband functions into centralized baseband units (BBUs). Which functions should stay with the remote radio head (RRH) and which ones should move to centralized BBUs has been extensively studied, and a few options exist.

The C-RAN model, which requires a fiber connection between the RRH and the BBUs, is popular in some regions of the world, such as China and Korea. This front-haul interface is challenging because it has distance limitations and is bandwidth limited. It uses a protocol known as the Common Public Radio Interface (CPRI). China Mobile has been in the forefront of the C-RAN architecture. Other countries such as Japan and Korea have also adopted C-RAN. In the U.S., Verizon is looking at implementing C-RAN for small cells.

Mobile Edge Computing (MEC)

Mobile Edge Computing (MEC) offers an important balance as cellular networks move to a central model. MEC allows certain latency-sensitive components to be moved to the network edge. For example, even though it may make business sense for MNOs to centralize several functions into VNFs running on standard off-the-shelf servers, some critical functions may need to be

closer to the edge of the network to ensure high availability, low latency, and higher levels of security.

TIP

The network edge is also a good place to host application servers that require local context. For example, if you are driving around and you receive restaurant ads on your mobile device based on your current location, you need the application service to be close to the edge to avoid a lengthy delay in the app that could render the information obsolete.

Chapter **6**

Creating a New 5G World Order with New Radio (NR)

n this chapter, you learn about a completely new technology that creates the foundation for 5G — the 5G New Radio (NR) physical air interface.

5G NR Basics

Unlike the other 5G elements (covered in previous chapters), that advance the industry toward the goals of 5G networks, 5G New Radio (NR) is not an evolution of 4G Long Term Evolution (LTE) innovations. 5G NR is a completely new technology specification — a physical air interface — that is required to achieve the extreme bandwidth, low latency, and massive scalability requirements of 5G.

REMEMBER

The other four areas of 4G LTE innovation that are being further developed for 5G include:

» **Speeds and feeds:** Discussed in Chapter 2, innovations such as carrier aggregation (CA), massive multiple input multiple output (MIMO), and quadrature amplitude modulation (QAM), among others, enables carriers to "fatten" the data pipe.

CHAPTER 6 **Creating a New 5G World Order with New Radio (NR)** 33

- » **Unlicensed spectrum:** Discussed in Chapter 3, LTE in unlicensed spectrum (LTE-U), License Assisted Access (LAA and enhanced LAA, or eLAA), and MulteFire enable carriers to leverage unlicensed spectrum as additional data pipes.

- » **Internet of Things (IoT):** Discussed in Chapter 4, IoT developments include NarrowBand IoT (NB-IoT), LTE for Machines (LTE-M or LTE Category M1), and Low Power Wide Area Network (LPWAN) specifications such as Long Range WAN (LoRaWAN) and Sigfox. These developments enable wireless communication with a diverse array of billions of IoT devices.

- » **Virtualization:** Discussed in Chapter 5, software-defined networking (SDN) and network functions virtualization (NFV) enable mobile network operators (MNOs) to achieve cost-effective scalability and elasticity in their core networks with innovations that include network slicing, virtual evolved packet core (vEPC), centralized radio access network (C-RAN), and mobile edge computing (MEC).

The 5G NR specification enables the following goals of 5G:

- » **Extreme bandwidths:** 5G NR aggregates eight component carriers (CCs). The carrier width specifications are still in development, but will likely be 100 megahertz (MHz) or greater, providing a total of approximately one gigahertz (GHz) of aggregated bandwidth — wide enough to carry 20 gigabits of data per second.

- » **Low latency:** In LTE Advanced, each subframe (there are 10 subframes) is handled in one millisecond (ms). 5G will be five to ten times faster — each subframe is handled in 100 to 200 microseconds (μs). 5G NR will use new channel coding techniques, such as low density parity check (LDPC), that are more efficient than existing techniques, resulting in shorter transmission time intervals (TTIs).

    A technique called *scalable orthogonal frequency-division multiplexing (OFDM) numerology* will be used to support different use cases and scenarios within the same frame. Effectively, shorter TTIs will be used for low latency, high reliability use cases, and longer TTIs for higher spectral efficiency, higher bandwidth use cases. 5G NR can also effectively multiplex between short and long TTIs, thereby allowing a diverse set of users to simultaneously use the system.

TECHNICAL STUFF

>> **Massive MIMO:** 5G NR extends MIMO up to 256 antenna elements and enables massive MIMO. This is a key enabler for higher spectrum bands. The antenna elements in this case are smaller, thereby making it less complex to build a massive array.

>> **Massive IoT:** 5G NR will use resource spread multiple access (RSMA) on the uplink to enable grant-free transmission of data on the uplink. A device does not need an enhanced Node B (eNodeB) to give it a grant (or slot) in the pipe to transmit data. This capability eliminates the need for signaling and allows devices to send small packets asynchronously. 5G NR will also address distance and location challenges in low-power IoT devices, using a technique called *multi-hop mesh* to relay uplink data via nearby devices.

5G NR is part of the Release 15 specification from the Third Generation Partnership Project (3GPP). Two versions of 5G NR exist:

>> **Non-standalone (NSA) 5G NR:** NSA will use the existing LTE radio and core network as the control plane anchor for mobility management and coverage, while adding a new 5G carrier. Early 2019 deployments of 5G NR will use this configuration. In this mode, the connection is anchored in 4G LTE while 5G NR carriers are used to boost data rates and reduce latency.

>> **Standalone (SA) 5G NR:** SA will use the new 5G core network (5GCN) architecture, including the full control and user plane offered by 5G.

An interim release of the 5G NR NSA specification was accelerated to the end of 2017 to enable large-scale trials and deployments for enhanced mobile broadband (eMBB) use cases to begin in 2019. The 5G NR SA specification is expected in mid-2018.

More Spectrum — mmWave Bands

In the U.S., the Federal Communications Commission (FCC) has opened a total of 10.85 GHz of spectrum above 24 GHz to enable 5G use cases. The new spectrum includes 3.85 GHz of licensed spectrum from 27.5 to 28.35 GHz and 37 to 40 GHz, and 7 GHz of unlicensed spectrum from 64 to 71 GHz.

These bands — between 30 GHz and 300 GHz — are known as millimeter wave bands (mmWave). These high frequency bands translate to narrow wavelengths in the range of one to ten millimeters. There are some challenges with mmWave, including:

» Short transmission paths and high propagation losses over long distances and anything that is not line of sight

» Weakened signals in gases and precipitation

However, the short transmission paths and propagation loss characteristics of mmWave enable spectrum reuse, by limiting the interference between adjacent cells. Additionally, because of the extremely short wavelengths of mmWave signals, transmission paths can be extended using small, multi-element dynamic beamforming antennas that can be installed in user equipment (UE) such as smartphones.

TIP

5G NR, operating at very high frequencies with small coverage areas (because of propagation and line-of-sight issues), is ideal for dense urban locations, where cell sizes are generally small (approximately 200 meters).

Chapter **7**

Exploring 5G Use Cases

I n this chapter, you learn about some of the key use cases that will be enabled by 5G.

Fixed Wireless Broadband Service

5G will begin its commercial drive with the fixed wireless broadband use case in the U.S. AT&T and Verizon are racing to be the first to deliver fixed wireless broadband service to their customers.

Verizon acquired XO Communications, which included the 28 gigahertz (GHz) spectrum, in 2016. One of the primary drivers for Verizon will be to offer 5G as an alternative to cable/advanced digital subscriber line (ADSL) for high-speed broadband access to residential customers for gigabit speed Internet, 4K content, virtual reality (VR), and more. The goal is to achieve greater than one gigabit per second (Gbps) speeds without digging trenches to get fiber to your home.

AT&T is also pursuing 5G on fixed wireless in both the 28 GHz and 37 GHz spectrum bands. 5G video trials are currently underway for some residential customers, allowing them to stream DIRECTV NOW video service over a fixed wireless 5G connection. In its internal testing, AT&T Wireless claims it can now achieve 13 Gbps over a 5G connection. Like Verizon, AT&T is also looking at 5G to replace home broadband connections delivered by cable companies.

If successful, these use case trials could signal the end of the "cable guy" coming out "sometime between 1 p.m. and 4 p.m." to route a coaxial cable through your home. In areas where there is currently only one option for home broadband or television service, this technology could also lead to more competition, better service, and lower rates for subscribers.

Cable companies, such as Charter Communications, are also exploring 5G technologies. Charter is planning to conduct 28 GHz 5G experiments in the near future with antennas mounted on a mobile trailer and van with hydraulic masts, which will be moved to each of the test locations.

TIP

The motivation for a company like Charter is that 5G will enable a host of new products that take advantage of low-latency, high-capacity networks, including virtual reality (VR) and augmented reality (AR) applications, many of which will be used in a fixed home or office location.

Entertainment Everywhere

The natural progression of the fixed wireless broadband use case (discussed in the previous section) enables mobile broadband at extreme data rates "everywhere" — not only to a fixed home or office location.

Although some of the envisioned mobile broadband experience started in Long Term Evolution (LTE) networks, 5G takes it to a different level with 4K (and subsequently 8K) television, high dynamic range video streaming, 3D videos, and more. All these applications require very high bandwidth, and many require always-on connectivity to push real-time information to the users. The transformation from fixed wireless to wireless everywhere

requires several mobility-specific factors to be addressed, such as fast handoffs, seamless connectivity at very high speeds and low latency, high reliability when mobile, and others.

REMEMBER

Cloud storage is also driving the growth of uplink data rates. In the past, content was mostly downloaded and thus required a "fat" pipe in only one direction: the downlink from the base station to the mobile device. However, content is increasingly being uploaded to different cloud storage platforms, requiring robust data pipes in both the downlink and uplink directions. Cloud gaming is another 5G driver that requires fast responsiveness and high broadband capacity.

Virtual Reality (VR) and Augmented Reality (AR)

Virtual reality (VR) technology creates a fully immersive, computer-generated experience that simulates or re-creates real-life situations and environments. In contrast to VR, augmented reality (AR) layers computer-generated images and enhancements onto a real-world situation or environment to provide a more meaningful context for user interaction.

Although current 4G networks are sufficient for some early-adopter VR and AR experiences, the introduction of 5G will enable more novel VR and AR experiences and make them available for mass adoption by consumers. Offering much more capacity, lower latency, and a more uniform experience, 5G will not only improve, but will also be a requirement for some of the most exciting AR and VR use cases, including:

» Sharing live streaming content on social media from event venues along with 50,000 other people in a stadium becomes even more challenging with 4K 360-degree video because each user is uploading 25 Mbps at the same time.

» Next-generation VR and AR experiences will have "six degrees of freedom" (6DoF) — the next level of immersion — allowing users to move within and intuitively interact with the environment. 6DoF content is an order of magnitude richer in naturalness and interactivity than current "three degrees of freedom" (3DoF) video. 3DoF experiences, such as 360-degree

video, allow the user to look around rotationally from a fixed position. 6DoF experiences, which are available in video games today, allow the user to move spatially through the environment just by walking or leaning their head forward.

6DoF head-motion tracking is required to enjoy 6DoF content in an intuitive manner. Many industries such as tourism, education, and other forms of immersive video will flourish as 6DoF technologies evolve. Most components of the video delivery pipeline are currently ill-suited for 6DoF video, including capture devices, production software, codecs, compression algorithms, the network, and players. 6DoF video also demands bit rates in the range of 200 Mbps to 1Gbps, depending on the end-to-end latency.

Connected and Autonomous Vehicles

Smart, connected cars are already here — and some security risks have already been notoriously demonstrated — and self-driving autonomous vehicles are beginning to appear on our roads. Other applications that 5G will enable for smart, connected cars include:

» **Traffic safety:** This includes the ability to detect hazardous road conditions, such as inclement weather or nearby accidents, and provide real-time guidance for appropriate courses of action to enable safer driving and reduce the risk of accidents. For example, imagine a situation in which your driverless car gets a real-time message (in microseconds) that a truck is rapidly approaching the intersection that you are about to cross. Your car then automatically slows down to let the truck pass the intersection, thereby avoiding a possible accident.

» **Entertainment (for passengers):** Live video and music streaming (including 4K ultra high-definition movies), interactive video games, cloud connectivity, and data exchange elaborate the need for high capacity and high mobility mobile broadband.

» **Augmented reality (AR):** Displaying key information in near real-time for drivers requires low latency so that the information is timely and relevant. In the earlier truck

example, a live situation is being tracked by the involved IoT devices and infrastructure, and relayed in real-time to the users — vehicle and human.

>> **Self-driving, autonomous vehicles:** These will require ultra-reliable, high-speed communication between different driverless cars, and between cars and infrastructure.

The Connected World

Smart homes, smart cities, multiple industries (such as health, retail, smart grids, and remote factories) have a common thread — they are using more and more devices and sensors that communicate with one another and the rest of the world.

Many of these devices are mission critical; others may send high-definition video, requiring high availability and very low latency. Yet another set of devices may send small data packets relatively infrequently (for example, every few hours, days, or weeks). Some examples of common use cases with diverse connectivity requirements include:

>> **Logistics and freight:** These devices and sensors typically require lower data rates, but need wide coverage and reliable location information.

>> **Smart grid:** A smart grid requires low latency sensors to regulate the use of utilities such as electricity, natural gas, and water. Leveraging digital information, such as the behaviors of suppliers and consumers, allows the smart grid to improve the efficiency, reliability, economics, and sustainability of the production and distribution of these resources.

>> **Remote medical:** Collaborating about a medical case with other surgeons located thousands of miles away was a use case scenario discussed as part of the Long Term Evolution (LTE) rollout. It can become a reality with the extreme bandwidth, low latency, and high availability of 5G networks.

>> **Hazardous areas:** The ability to remotely explore mining areas or shut down a nuclear power plant during an emergency — in a fraction of the time required for human interaction, and without risk to human life — is possible with 5G.

Augmented Humans

Looking ahead to the more distant future — perhaps 30 years from now — Google futurist Ray Kurzweil predicts humans will be able to upload their entire minds to computers and become "digitally immortal" — an event called *singularity.*

In 2011, IBM's Watson beat former winners on the television game show *Jeopardy!,* proving that computers can outperform the best of humans when it comes to synthesizing information and beating them to the buzzer. Google Home and Amazon Echo are becoming more common in homes — perhaps a bit unnerving at first as they start entering our private lives, and increasingly taking center stage when we want them to tell us more about the weather, play our favorite song, read us our favorite *5G For Dummies* book, or tell a few jokes. As machine learning and artificial intelligence develop further, these intelligent devices will better understand human behavior and evolve beyond databases of information. Eventually, they will also become more mobile and intrinsic in our lives, rather than sitting on a shelf in our living rooms.

REMEMBER

5G — with its extreme bandwidths, very low latency, and massive scale support — will be a critical component in creating these real-time experiences, initially as an assistant to humans, and eventually even "thinking" for humans (in some cases). Shopping for clothes with your kids or selecting your next car will become a more immersive experience. 5G technology might also assist humans as an augmented partner, offloading some of the tasks humans don't want to "expend their own brain cells" on. Although it may sound like science fiction, 5G might be a starting point for the next evolution of humans and machines.

IN THIS CHAPTER

» **Going beyond speed and latency**

» **Looking at drivers and timelines**

» **Overcoming misconceptions about limitations**

» **Recognizing complementary technologies**

» **Transforming entire industries**

Chapter **8**

Ten Myths About 5G — Debunked

n this chapter, we expose ten common myths about 5G — and we clue you in to the reality behind them. Keep these points in mind:

» **5G is all about higher speeds to the user.** Although one of the key goals of 5G is to provide extreme bandwidth (high-speed data) to users, low latency and massive scale are other key goals of 5G. So, 5G isn't just about speed!

» **5G requires less than one millisecond latency.** Although less than one millisecond of latency is a goal of 5G, 5G networks will be deployed before that target is actually achieved.

» **Smartphones will lead the charge to 5G.** The iPhone and Android were born in the 3G era and virtually exploded (literally, in some cases) during the 4G LTE era. However, 5G will not only enable faster and better smartphones — it will also lead to mass-market consumer VR and AR devices, sensors and applications for smart homes and cars, industrial robots, and billions of other Internet of Things (IoT) devices yet to be conceived.

»» 5G will be commercially available in 2017. Although Verizon plans to roll out 5G fixed mobility in 2017 and Olympic 5G trials will be underway during the 2018 Winter Olympics, 5G standards will not be finalized until the end of 2017. Thus, although some early, pre-standard versions of 5G will be coming soon, the 5G standard is not complete. Standards-compliant 5G will be coming much later.

»» 5G is only for short-range, line-of-sight communication. In addition to other frequency bands, 5G uses mmWave bands, which are ideal for very short ranges. However, plenty of ongoing experiments demonstrate how techniques such as beamforming can achieve greater ranges to users in challenging environments beyond line-of-sight.

»» 5G will be used only in very high bands. Although 5G will be deployed in very high millimeter wave (mmWave) bands, it will also re-use spectrum in lower bands, both licensed and unlicensed.

»» 5G will replace 4G LTE. 5G will coexist with 4G LTE for a long time to come. 4G has plenty to offer for many current applications such as voice, data, and even IoT.

»» 5G will be a revolution, not an evolution. Although 5G brings in a new physical layer (in 5G New Radio, or 5G NR), there is plenty of evolution from LTE-A Pro technologies such as carrier aggregation (CA), massive multiple input/multiple output (MIMO), quadrature amplitude modulation (QAM), unlicensed spectrum (LTE in unlicensed spectrum or LTE-U, License Assisted Access or LAA, and MulteFire, among others), IoT, and virtualization.

»» 5G will be required to drive IoT. IoT will initially be driven by LTE-A Pro where NarrowBand IoT (NB-IoT) is specified. In addition, other low-power technologies, such as Long Range Wide Area Network (LoRaWAN) and Sigfox, have been defined for IoT.

»» The 5G winners will be the operators and vendors. Mobile network operators (MNOs), network equipment manufacturers (NEMs), and smartphone manufacturers were the primary business beneficiaries of 4G LTE. However, 5G will transform many industries, including car manufacturing, agriculture, health and medicine, transportation and logistics, and many more.

5 Keys to 5G

1. **5G-NR (New Radio)** — Next-Gen XAir
2. **5G Speeds/Feeds** — Ixia Multi-Gig LTE with XAir2
3. **Wi-Fi Coexistence** — Ixia LTE Unlicensed
4. **Massive IoT** — Ixia IoT
5. **Testing NFV Deployments** — IxLoad Wireless VE

ixia

Read more about about enabling the 5G revolution at www.ixiacom.com/solutions/5g-wireless-test

Dream up new opportunities in the 5G future

The next-generation mobile network is on the horizon. 5G, the successor to 4G LTE networks, will enable significantly greater mobile speeds — as much as 20 gigabits per second — with very low latency. 5G will support a massive array of devices, from smart phones to virtual reality devices to small sensors. In this book, you learn about the technological innovations being developed today to enable a 5G future. You also learn about potential use cases that will transform entire industries and create new business models and opportunities.

Inside…

- Where wireless communication is headed
- Why higher speeds are the future
- How the IoT will employ 5G connectivity
- Why the unlicensed spectrum matters
- Why New Radio is a starting point for 5G
- How 5G will enable new business models

ixia

Kalyan Sundhar is Vice President — Mobility, Virtualization, and Applications at Ixia. He has more than 25 years experience in the industry, building cutting-edge products. **Lawrence C. Miller** has worked in information technology for more than 25 years. He has written more than 60 *For Dummies* books.

Go to Dummies.com®
for videos, step-by-step photos, how-to articles, or to shop!

for dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.

LTE to 5G: Cellular and Broadband Innovation

Table of Contents

Introduction

Mobile computing with wireless communications has already changed how people socialize and how companies do business. Yet, we are still in the nascent stages of the transformation that ubiquitous connectivity is enabling. Early examples of this exciting future include virtual and augmented reality, autonomous driving, smart cities, wearable computers, and connected devices throughout our environment. While no one can predict the full extent of innovation that the new global broadband fabric, aided by complementary innovations such as AI, will unleash, one thing is certain: We are rushing toward an extraordinary time.

The step from 3G to 4G was dramatic, and industry advances occurring now in LTE and 5G will be even greater. Standards bodies are standardizing 5G, a process that will continue through the 2020 timeframe, with ongoing enhancements continuing during the next decade. Some operators are deploying pre-standard 5G in trials this year, and initial standards-based networks could be deployed as early as 2019. 5G will not replace LTE; in most deployments, the two technologies will be tightly integrated and co-exist through at least the late-2020s.

Many of the capabilities that will make 5G so effective are appearing in advanced forms of LTE. With carrier aggregation, for example, operators have not only harnessed the potential of their spectrum holdings to augment capacity and performance, but the technology is also the foundation for entirely new capabilities, such as operating LTE in unlicensed bands.

The computing power of today's handheld computers rivals that of past mainframe computers, powering intuitive operating systems and millions of applications. Coupled with affordable mobile broadband connectivity, these devices provide such unprecedented utility that billions of people are using them.

With long term growth in smartphone and other mobile device use limited by population, innovators are increasingly turning their attention to the Internet of Things (IoT), which already encompasses a wide array of applications. Enhancements to LTE, followed by 5G IoT capabilities, will connect wearable computers, sensors, and other devices, leading to better health, economic gains, and other advantages. 5G addresses not only IoT deployments on a massive scale but also enables applications previously not possible that depend on ultra-reliable and low-latency communications, sometimes called "mission-critical applications." Although a far more fragmented market than smartphones, the benefits will be so great that the realization of IoT on a massive scale is inevitable.

Regulatory policies are striving to keep pace, addressing complex issues that include how best to allocate and manage new spectrum, network neutrality, and privacy. Policy decisions will have a major impact on the evolution of mobile broadband.

These are exciting times for both people working in the industry and those who use the technology—which today is nearly everybody. This paper attempts to capture the scope of what the industry is developing, beginning with Table 1, which summarizes some of the most important advances.

Table 1: Most Important Wireless Industry Developments in 2017

Development	Summary
5G Research and Development Accelerates	5G, in early stages of definition through global efforts and many proposed technical approaches, could be deployed in non-standalone versions as early as 2019. Deployment will continue through 2030. Some operators will deploy pre-standard networks for fixed-wireless access in 2017. 5G is being designed to integrate with LTE, and some 5G features may be implemented as LTE-Advanced Pro extensions prior to full 5G availability.
5G New Radio (NR) Being Defined	Key aspects of the 5G NR have been determined, such as radio channel widths and use of OFDMA. The first version, specified in Release 15, will support low-latency, beam-based channels, massive Multiple Input Multiple Output (MIMO) with large numbers of controllable antenna elements, scalable-width subchannels, carrier aggregation, Cloud Radio-Access Network (RAN) capability, and dynamic co-existence with LTE.
LTE Has Become the Global Cellular Standard	A previously fragmented wireless industry has consolidated globally on LTE. LTE is being deployed more quickly than any previous-generation wireless technology.
LTE-Advanced Provides Dramatic Advantages	LTE capabilities continue to improve with carrier aggregation, 1 Gbps peak throughputs, higher-order MIMO, multiple methods for expanding capacity in unlicensed spectrum, new IoT capabilities, vehicle-based communications, small-cell support including Enhanced Inter-Cell Interference Coordination (eICIC), lower latency, Self-Organizing Network (SON) capabilities and Enhanced Coordinated Multi Point (eCoMP).
Internet of Things Poised for Massive Adoption	IoT, evolving from machine-to-machine (M2M) communications, is seeing rapid adoption, with tens of billions of new connected devices expected over the next decade. Drivers include improved LTE support, such as low-cost and low-power modems, enhanced coverage, higher capacity, and service-layer standardization, such as oneM2M.
Unlicensed Spectrum Becomes More Tightly Integrated with Cellular	The industry has developed increasingly sophisticated means for integrating Wi-Fi and cellular networks, such as LTE-WLAN Aggregation (LWA) and LTE-WLAN Aggregation with IPSec Tunnel (LWIP), making the user experience ever more seamless. The industry has also developed and is now deploying versions of LTE that can operate in unlicensed spectrum, such as LTE-Unlicensed (LTE-U), LTE-Licensed Assisted Access (LTE-LAA), and

Development	Summary
	MulteFire. Cellular and Wi-Fi industry members are successfully collaborating to ensure fair spectrum co-existence.
Spectrum Still Precious	Spectrum in general, and in particular licensed spectrum, remains a precious commodity for the industry.
	Recently added spectrum in the United States includes the 600 MHz band, auctioned in 2017, and the 3.5 GHz Citizens Broadband Radio Service (CBRS) "small-cell" band, which could see initial deployments in 2018.
	5G spectrum will include bands above 6 GHz, including "mmWave" (30 GHz to 100 GHz), with the potential of ten times (or more) as much spectrum as is currently available for cellular. Radio channels of 200 MHz and 400 MHz, and even wider in the future, will enable multi-Gbps peak throughput.
Small Cells Take Baby Steps, Preparing to Stride	Operators have begun installing small cells, which now number in the tens of thousands. Eventually, hundreds of thousands if not millions of small cells will lead to massive increases in capacity.
	The industry is slowly overcoming challenges that include restrictive regulations, site acquisition, self-organization, interference management, power, and backhaul.
Network Function Virtualization (NFV) Emerges	Network function virtualization (NFV) and software-defined networking (SDN) tools and architectures are enabling operators to reduce network costs, simplify deployment of new services, reduce deployment time, and scale their networks.
	Some operators are also virtualizing the radio-access network, as well as pursuing a related development called cloud radio-access network (cloud RAN). NFV and cloud RAN will be integral components of 5G.

The main part of this paper covers the transformation of broadband, exploding demand for wireless services, the path to 5G including planned and expected capabilities, new LTE innovations, supporting technologies and architectures, voice over LTE (VoLTE), Wi-Fi Calling, LTE for public safety, options to expand capacity, and spectrum developments.

The appendix delves into more technical aspects of the following topics: 3GPP Releases, Data Throughput, latency, 5G, LTE, LTE-Advanced, LTE-Advanced Pro, HetNets and small cells, IoT, cloud RANs, Unlicensed Spectrum Integration, self-organizing networks, the IP Multimedia Subsystem, broadcast/multicast services, backhaul, UMTS/WCDMA,[1] HSPA, HSPA+, UMTS TDD, and EDGE/EGPRS.

---

[1] Although many use the terms "UMTS" and "WCDMA" interchangeably, in this paper "WCDMA" refers to the radio interface technology used within UMTS, and "UMTS" refers to the complete system. HSPA is an enhancement to WCDMA.

Broadband Transformation

Broadband networks rely on a fiber core with various access technologies, such as fiber to the premises, coaxial cable, digital subscriber line (DSL), or wireless connections. LTE provides a broadband experience, but capacity limitations prevent it from being the only broadband connection for most users. As a result, a majority of consumers in developed countries have both mobile broadband and fixed broadband accounts.

Two developments will transform the current situation:

1. **Fiber Densification**. Multiple companies are investing to extend the reach of fiber, decreasing the distance from the fiber network to the end node.

2. **5G Standardization and Deployment.** As 5G mmWave technology, including massive MIMO and beamforming, becomes commoditized, it will increasingly be a viable alternative to fixed-access technologies such as coaxial, DSL, and even fiber connections. 5G commercial services will enable a new innovation cycle. The ability to create new applications and services with fewer limitations will take the connected society to a new level.

Consequently, the companies that provide broadband service may change, and eventually, fixed and mobile broadband services may converge. For a more detailed discussion of trends in broadband, including the disruptive role of mmWave, refer to the 2017 Datacomm Research and Rysavy Research report, *Broadband Disruption: How 5G Will Reshape the Competitive Landscape.*[2]

As shown in Figure 1, the emerging broadband network is one with denser fiber and competing access technologies in which wireless connectivity plays a larger role.

---

[2] Details at https://datacommresearch.com/reports-broadband/.

Figure 1: Fiber Densification with Multiple Access Technologies, including mmWave

Many elements are interacting to transform mobile broadband, but the factors playing the most important roles are emerging capabilities for IoT, radio advances granting access to far more spectrum, small cells, new network architectures that leverage network function virtualization and software-defined networking, and new means to employ unlicensed spectrum. Except for access to high-band spectrum, a 5G objective, these advances apply to both LTE and 5G.
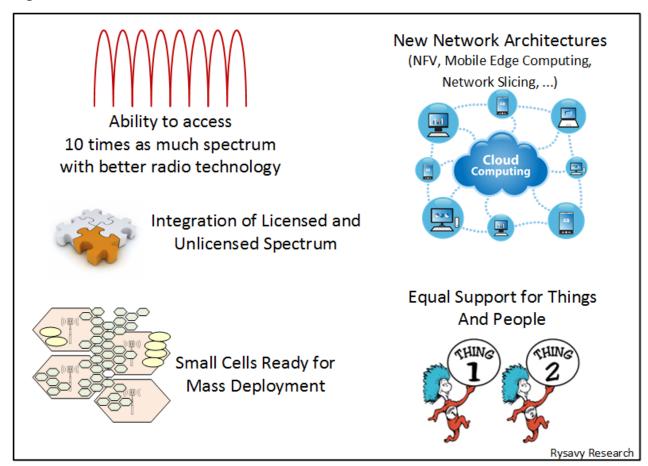
Figure 2: Fundamental Mobile Broadband Transformational Elements

In the past, developers used modems and networks designed for human communication for machine-type applications. But now, new modes of network operation, initially in LTE then enhanced further in 5G, will cater to the unique needs of a wide variety of machine applications, addressing low-cost, long battery life, long communications range, and a wide variety of throughputs. For instance, some IoT applications need only low-throughput communications, some sending only a small number of bits per day.

As for spectrum, throughout radio history, technology has climbed up a ladder to use higher frequencies. What were called "ultra-high frequencies" when made available for television are now considered low-band frequencies for cellular. Frequencies above 6 GHz, particularly mmWave frequencies, are the new frontier. Networks will ultimately take advantage of ten times as much spectrum as they use now, and likely even more over time. Although challenging to use because of propagation limitations and higher penetration loss, massive MIMO, beam steering, beam tracking, dual connectivity, carrier aggregation, small-cell architectures with self-backhauling, and other methods will help mitigate the challenges at these frequencies. The result: massive increases in capacity in localized areas.

In addition to accessing higher bands, cellular technologies are integrating unlicensed spectrum more efficiently, using technologies such as LTE-U, LAA, MulteFire, LWA, and LWIP. This integration will immediately augment small-cell capacity, improving the business case for small cells.

Small cells, on the roadmap for many years but held back by implementation difficulties, such as backhaul, are on the verge of large scale deployment, leading ultimately to densities as high as ten small cells or more for every macro cell. Paving the way are better wireless backhaul solutions, neutral-host capabilities enabled by new technologies, and soon, access to mmWave bands.

Facilitating the capabilities listed above, networks are becoming programmable. Using a distributed, software-enabled network based on virtualization and new architectural approaches such as Multi-access Edge Computing (MEC) and network slicing, operators and third parties will be able to deploy new services and applications more rapidly, and in a more scalable fashion. Centralizing RAN signal processing will also play a huge role; depending on the deployment scenario, such centralization will increase RAN efficiency and decrease deployment cost.

This paper lists the dozens of other innovations also fueling mobile and cellular technology transformation. Together, these transformed networks will mean that for millions, and ultimately billions, of people, wireless connections will be the only connections that they need. These networks will also provide the foundation for entire new industries, ones not yet even conceived.

Exploding Demand

Two technology trajectories have created critical mass: handheld computing and fast wireless connections. This combined computing and communications platform inspires the innovation that has produced millions of mobile applications.

IoT is a third trajectory. LTE, and eventually 5G, will connect tens of billions of devices. And fixed-wireless access could be a fourth. 5G, with expected network capacity a hundred times greater than 4G networks, will make wireless connections not only a viable substitute for wireline broadband connections for tens of millions of users over the next ten years, but make that broadband connection the only connection that many users need.

Figure 3: Exploding Demand from Critical Mass of Factors

This section analyzes some of these demand factors.

Application Innovation

When planning 4G network technology, who could have predicted applications such as Uber and Lyft, which combine location information with mapping and online payment and are now disrupting the taxi industry and even challenging notions of private vehicle ownership? While some applications of new technology can be predicted, many cannot.

More efficient technology not only addresses escalating demand, it also provides higher performance, thus encouraging new usage models and further increasing demand.

Today's smartphones and tablets, dominated by the iOS and Android ecosystems, in combination with sophisticated cloud-based services, provide a stable, well-defined application environment, allowing developers to target billions of users. Developers have rich platform-specific development tools; web-based tools such as HTML5; application programming interfaces (APIs) for mobile-specific functions, such as WebRTC (Web Real-Time Communications); and cloud-based services for applications and application services, such as notifications, IoT support, and mobile-commerce.

Of concern to some companies in the wireless industry, however, are current network neutrality rules that could possibly hamper innovation. By restricting prioritization, for example, the rules seem to fail to recognize that traffic streams from different applications have inherently different quality of service requirements.[3]

Internet of Things

Current M2M and Internet of Things applications include vehicle infotainment, connected healthcare, transportation and logistics, connected cars, home security and automation, manufacturing, construction and heavy equipment, energy management, video surveillance, environmental monitoring, smart buildings, wearable computing, object tracking, and digital signage. Municipalities, evaluating the concept of "smart cities," are exploring how to optimize pedestrian and vehicular traffic, connect utility meters, and deploy trash containers that can report when they need emptying.

Although promising, the IoT market is also challenging, with varying communications requirements, long installation lifetimes, power demands that challenge current battery technology, cost sensitivity, security and data privacy concerns, unsuitability of conventional networking protocols for some applications, and other factors that developers must address. Consequently, the IoT opportunity is not uniform; it will eventually comprise thousands of markets. Success will occur one sector at a time, with advances in one area providing building blocks for the next.

Cloud-based support platforms and standardized interfaces are essential for development and deployment of IoT applications. For example, oneM2M has developed a service-layer that can be embedded in hardware and software to simplify communications with application servers.[4]

To address the IoT opportunity, 3GPP is defining progressive LTE refinements that will occur over multiple 3GPP releases. These refinements include low-cost modules that approach 2G module pricing and enable multi-year battery life. See the section "Internet of Things and Machine-to-Machine" in the appendix for more details.

Video Streaming

Video represents the greatest usage of data on smartphones. Just an hour a day of mobile video at 1.0 Mbps throughput, common with YouTube or Netflix, consumes 13.5 GB per month. Many streams, now available in high-definition (HD), consume even more data.

---

[3] For further discussion, see Rysavy Research, *How "Title II" Net Neutrality Undermines 5G*, April 19, 2017. Available at http://www.rysavy.com/Articles/2017-04-How-Title-II-Net-Neutrality-Undermines-5G.pdf. Also see *How Wireless is Different – Considerations for the Open Internet Rulemaking*, September 12, 2014. Available at http://www.rysavy.com/Articles/2014-09-Wireless-Open-Internet.pdf.

[4] OneM2M home page: http://onem2m.org/.

See the Appendix section "Data Consumed by Video" for a quantification of data consumed by video for multiple usage scenarios.

An increasing number of video applications adapt their streaming rates based on available bandwidth. By doing so, they can continue to operate even when throughput rates drop. Conversely, they take advantage of higher available bandwidth to present video at higher resolution. Fortunately, application developers are becoming sensitive to bandwidth constraints and are offering options for users to reduce consumption.

Nevertheless, video can consume so much data that cutting the broadband cord at the same time as the television cord is not possible for most consumers. With 5G's massive increase in capacity, however, an increasing percentage of users will use wireless connections as their only broadband connections, including for all their entertainment needs.

Cloud Computing

Cloud computing inherently increases data consumption because it requires communications for all operations. Examples include data synchronization and backup, cloud-based applications (such as email, word processing, and spreadsheets), automatic photo uploads, and music and video streaming.

5G Data Drivers

Futurists can predict some 5G applications, but many others will arise as industries evolve or come into existence to take advantage of new network capabilities. Some potential applications of 5G include:

- ❑ Ultra-high-definition, such as 4K and 8K, and 3D video.

- ❑ Augmented and immersive virtual reality. Ultra-high-fidelity virtual reality can consume 50 times the bandwidth of a high-definition video stream.

- ❑ Realization of the tactile internet—real-time, immediate sensing and control, enabling a vast array of new applications.

- ❑ Automotive, including autonomous vehicles, driver-assistance systems, vehicular internet, infotainment, inter-vehicle information exchange, and vehicle pre-crash sensing and mitigation.

- ❑ Monitoring of critical infrastructure, such as transmission lines, using long battery life and low-latency sensors.

- ❑ Smart transportation using data from vehicles, road sensors, and cameras to optimize traffic flow.

- ❑ Mobile health and telemedicine systems that rely on ready availability of high-resolution and detailed medical records, imaging, and diagnostic video.

- ❑ Public safety, including broadband data and mission-critical voice.

- ❑ Sports and fitness enhancement through biometric sensing, real-time monitoring, and data analysis.

- ❑ Fixed broadband replacement.

Some of these applications are already being addressed by 4G, but 5G's lower costs, higher throughputs, and lower latency will hasten realization of their potential.

Specific industries expected to take advantage of 5G include manufacturing, healthcare, media and entertainment, financial services, public safety, automotive, public transport, and energy utilities.[5]

Global Mobile Adoption

Figure 4 shows the often-cited Cisco projection of global mobile data consumption through 2021, measured in exabytes (billion gigabytes) per month, demonstrating traffic growing at a compound annual rate of 47%.

Figure 4: Global Mobile Data 2016 to 2021[6]
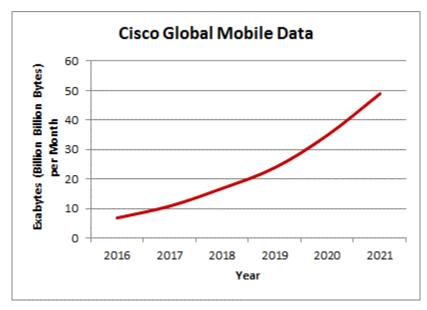
Figure 5 shows another data projection, predicting 45% annual growth in data for the 2016-to-2022 period, resulting in eight-fold growth.

---

[5] For further insight, refer to the Ericsson white paper, *The 5G Business Potential,* February 2017. Available at https://www.ericsson.com/en/networks/insights/the-5g-business-potential.

[6] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021*, February 2017.
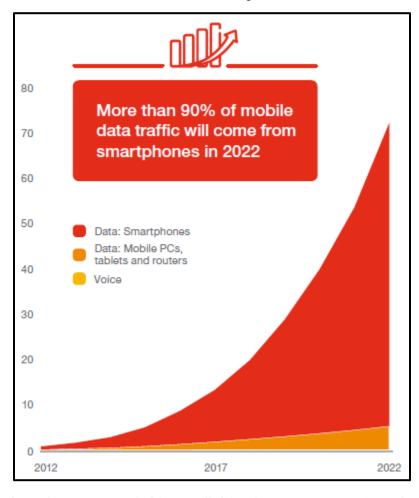
Figure 5: Global Mobile Voice and Data (Exabytes/Month) 2012 to 2022[7]

Figure 6 shows how the Internet of Things will drive far more connections than are launched by people.

---

[7] Ericsson, *Ericsson Mobility Report*, June 2017.
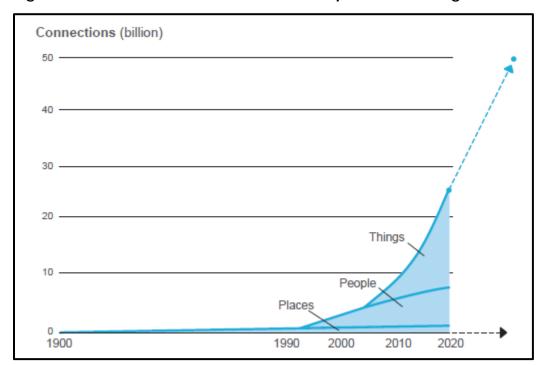
Figure 6: Connections of Places Versus People Versus Things[8]

In June 2017, more than 7.56 billion GSM-HSPA-LTE connections were in effect[9]—greater than the world's 7.40 billion population.[10] By the end of 2022, the global mobile broadband market is expected to include nearly 9.4 billion subscribers, with 9.3 billion using 3GPP technologies, representing more than 99% market share.[11]

LTE has experienced faster deployment than any mobile technology ever developed. All major U.S. operators now offer nationwide LTE coverage. LTE has also been chosen by U.S. national public-safety organizations as their broadband technology of choice.

As shown in Figure 7, 2G GSM has peaked and is now declining, as is CDMA. Both HSPA and LTE subscriptions will continue to rise through the rest of the decade.

---

[8] Ericsson, *IoT Security*, February 2017, available at
https://www.ericsson.com/assets/local/publications/white-papers/wp-iot-security-february-2017.pdf.

[9] Ovum, July 2017.

[10] U.S. Census Bureau, "U.S. and World Population Clock," http://www.census.gov/popclock/, accessed July 18, 2017.

[11] Ovum, July 2017. Note that the 2018 mobile broadband market figures include GSM/EDGE, because most GSM networks are likely to include Evolved EDGE, which provides mobile broadband capability.
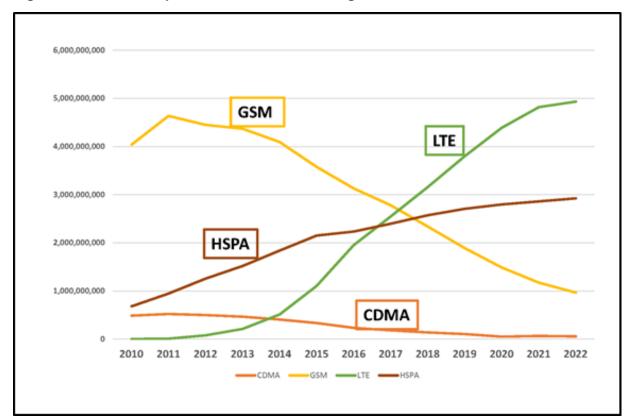
Figure 7: Global Adoption of 2G-4G Technologies 2010 to 2020[12]

IoT is growing rapidly as well, although connections so far are well below phone-type connections. Cisco projects 3.3 billion IoT connections by 2021, with 6% on 2G cellular, 16% on 3G, and 46% on 4G.[13]

---

[12] Ovum (WCIS), July 2017.

[13] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021*, February 2017.

Almost at 5G

With intensive development occurring within 3GPP, the first 5G specification will be completed in early 2018, enabling the first standards-based networks to be deployed in the 2019-2020 timeframe. This section includes expanding use cases, 1G-to-5G evolution, 4G LTE advances, 5G radio and architectures, Information-Centric Networking, phasing of 5G releases, and an overview of 3GPP releases.

Expanding Use Cases

Many wireless technology discussions focus on radio capabilities, but other equally important aspects include use cases, the services built on top of the technology, how different networks integrate with one another, and the topology of the networks. As summarized in Figure 8, these aspects are expanding, making mobile/wireless technology the foundation for other enterprises, including business-process optimization, consumer electronics, M2M, connected devices, and a multitude of vertical industries.

Figure 8: Expanding Use Cases

Table 2 summarizes the requirements of the expanding number of use cases that employ wireless technology.

Table 2: Requirements for Different Use Cases[14]

Use Cases	Requirements	Desired Value
Autonomous vehicle control	Latency	5 msec
	Availability	99.999 percent
	Reliability	99.999 percent

---

[14] Ericsson, *5G Systems – Enabling the Transformation of Industry and Society,* January 2017. Available at https://www.ericsson.com/assets/local/publications/white-papers/wp-5g-systems.pdf. Adapted from Table 1.

Use Cases	Requirements	Desired Value
Emergency communication	Availability	99.9 percent victim discovery rate
	Energy efficiency	One-week battery life
Factory cell automation	Latency	Down to below 1ms
	Reliability	Down to packet loss of less than 10^{-9}
High-speed train	Traffic density	Downlink (DL): 100Gbps/km^2, uplink (UL): 50 Gbps/km^2
	User throughput	DL: 50Mbps, UL: 25Mbps
	Mobility	500 km/h
	Latency	10ms
Large outdoor event	User throughput	30Mbps
	Traffic density	900Gbps/km^2
	Connection density	Four devices/m^2
Massive IoT	Connection density	1,000,000 devices/km^2
	Availability	99.9 percent coverage
	Energy efficiency	10-year battery life
Remote surgery and examination	Latency	Down to 1ms
	Reliability	99.999 percent
Smart city	User throughput	DL: 300Mbps, UL: 60Mbps
	Traffic density	700 Gbps/km^2
	Connection density	200,000 devices/km^2
Virtual and augmented reality	User throughput	4-28Gbps
	Latency	< 7msec
Broadband to the home	Connection density	4,000 devices/km^2
	Traffic density	60Gbps/km^2

The economic role that wireless technology plays keeps increasing. One study anticipates that in 2035, 5G will enable $12.3 trillion of global economic output.[15]

1G to 5G Evolution

5G is almost upon us; it is a technology that will be constructed from millions of ideas, methods, algorithms, and processes. Just as 4G LTE became available when previous technologies, such as HSPA, could be further improved, 5G enters the stage when the roadmap for LTE has not been exhausted. And just as 2G coexists today with 3G and 4G, 5G will co-exist with previous generations of technology.

For historical context, "1G" refers to analog cellular technologies that became available in the 1980s. "2G" denotes initial digital systems that became available in the 1990s and that introduced services such as short messaging and lower-speed data. 3G requirements were specified by the International Telecommunication Union (ITU) as part of the International Mobile Telephone 2000 (IMT-2000) project, for which significant voice capacity improvement was a focus and digital networks had to provide 144 Kbps of throughput at mobile speeds, 384 Kbps at pedestrian speeds, and 2 Mbps in indoor environments. UMTS-

---

[15] HIS Economics and HIS Technology, *The 5G Economy: How 5G Technology will contribute to the global economy,* January 2017. Commissioned by Qualcomm Technologies.

HSPA and CDMA2000 are the primary 3G technologies. 3G technologies began to be deployed early last decade.

In 2008, the ITU issued requirements for IMT-Advanced, which many people initially used as a definition of 4G. The focus on 4G was to improve data coverage, capacity, and quality of experience. Requirements included operation in up to-40 MHz radio channels and extremely high Spectral Efficiency. The ITU required peak spectral efficiency of 15 bps/Hz and recommended operation in up-to-100 MHz radio channels, resulting in a theoretical throughput rate of 1.5 Gbps. In 2009 and 2010, the term "4G" became associated with mobile broadband technologies deployed at the time, such as HSPA+, WiMAX, and initial LTE deployments. Today, 4G usually refers to HSPA+ or LTE.

Although the industry is preparing for 5G, LTE capabilities will continue to improve in LTE-Advanced Pro through the rest of the decade. Many of these enhancements will come through incremental network investments. Given the scope of global wireless infrastructure, measured in hundreds of billions of dollars, offering users the most affordable service requires operators to leverage investments they have already made. 5G will eventually play an important role, but it must be timed appropriately so that the jump in capability justifies the new investment.

5G groups researching next generation wireless architecture and requirements include, among others, the International Telecommunication Union (ITU),[16] the European Union's 5G Infrastructure Public-Private-Partnership (5G PPP), which is the framework for several projects, including METIS II (Mobile and wireless communications Enablers for the Twenty-twenty Information Society), and Next Generation Mobile Networks (NGMN). Finally, 5G Americas is actively involved in developing the vision and requirements of 5G for North, Central, and South America. 5G Americas signed an MoU to collaborate with 5G-PPP.[17]

The ITU, the standardization group of the United Nations, has set the following standardization timetable in its IMT-2020 project:[18]

- ❑ 2016-2017: Definition of technical performance requirements, evaluation criteria and methods, and submission templates.

- ❑ 2018-2019: Submission of proposals.

- ❑ 2019: Evaluation of proposed technologies.

- ❑ 2019 and 2020: Publication of IMT-2020 specifications.

---

[16] International Telecommunication Union, "Working Party 5D (WP 5D) - IMT Systems," http://www.itu.int/ITU-R/index.asp?category=study-groups&rlink=rwp5d&lang=en, accessed May 16, 2017.

[17] 5GPPP, "5G-PPP MoU with 4G Americas," March 2, 2015. http://5g-ppp.eu/5g-ppp-mou-with-4g-amerericas/

[18] ITU, "ITU towards 'IMT for 2020 and beyond'," http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx, viewed May 16, 2017.

Wireless technology has progressed to the extent that significant new capabilities are inevitable, making 5G a possible alternative to wireline broadband for many subscribers.[19]

Table 3 summarizes the generations of wireless technology.

Table 3: 1G to 5G

Generation	Requirements	Comments
1G	No official requirements. Analog technology.	Deployed in the 1980s.
2G	No official requirements. Digital technology.	First digital systems. Deployed in the 1990s. New services such as SMS and low-rate data. Primary technologies include IS-95 CDMA (cdmaOne), IS-136 (D-AMPS), and GSM.
3G	ITU's IMT-2000 required 144 Kbps mobile, 384 Kbps pedestrian, 2 Mbps indoors.	First deployment in 2000. Primary technologies include CDMA2000 1X/EV-DO and UMTS-HSPA. WiMAX.
4G (Initial Technical Designation)	ITU's IMT-Advanced requirements include the ability to operate in up-to-40-MHz radio channels and with very high spectral efficiency.	First deployment in 2010. IEEE 802.16m and LTE-Advanced meet the requirements.
4G (Current Marketing Designation)	Systems that significantly exceed the performance of initial 3G networks. No quantitative requirements.	Today's HSPA+, LTE, and WiMAX networks meet this requirement.
5G	ITU IMT-2020 has defined technical requirements for 5G, and 3GPP is developing specifications.	First standards-based deployments in 2019 and 2020.

The interval between each significant technology platform has been about ten years. Within each platform, however, innovation surges unabated. For example, with 2G technology,

---

[19] Rysavy Research, "How will 5G compare to fiber, cable or DSL?" Fierce Wireless, May 2014. Available at http://www.rysavy.com/Articles/2014-05-5G-Comparison-Wireline.pdf.

EDGE significantly improved data performance compared with initial General Packet Radio Service (GPRS) capabilities. Similarly, HSPA hugely increased data speeds compared with initial 3G capabilities. LTE and LTE-Advanced are also acquiring continual improvements that include faster speeds, greater efficiency, and the ability to aggregate spectrum more flexibly.

Figure 9 presents the timeline of technology generations, including past and future, showing initial deployment, the year of the peak number of subscribers, and decline. Each cellular generation spans multiple decades, with peak adoption occurring some 20 years after initial deployment. 6G deployment in 2030, though highly speculative, is consistent with deployment of previous generations.
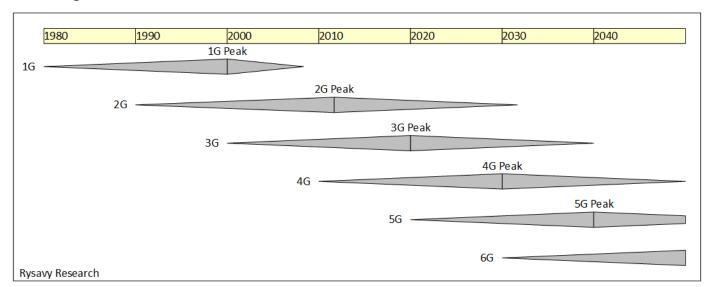
Figure 9: Timeline of Cellular Generations

4G LTE Advances

As competitive pressures in the mobile broadband market intensified, and as demand for capacity persistently grew, LTE became the favored 4G solution because of its high data throughputs, low-latency, and high spectral efficiency. Specifically:

❑ **Wider Radio Channels.** LTE can be deployed in wide radio channels (for example, 10 MHz or 20 MHz) with carrier aggregation now up to 640 MHz, although inter-band aggregation of four or five carriers (up to 100 MHz) represents a practical upper limit.

❑ **Easiest MIMO Deployment.** By using new radios and antennas, LTE facilitates MIMO Deployment, in contrast to the logistical challenges of adding antennas for MIMO to existing legacy technologies. Furthermore, MIMO gains are maximized because all user equipment supports it from the beginning.

❑ **Best Latency Performance.** For some applications, low latency (packet traversal delay) is as important as high throughput. With a low transmission time interval (TTI) of 1 millisecond (msec) and a flat architecture (fewer nodes in the core network), LTE has the lowest latency of any cellular technology.

In the same way that 3G coexists with 2G systems in integrated networks, LTE systems co-exist with both 3G and 2G systems, with devices capable of 2G, 3G, and 4G modes.

Beyond radio technology, the Evolved Packet Core (EPC) provides a core architecture that integrates with both legacy GSM-HSPA networks and other wireless technologies, such as CDMA2000 and Wi-Fi. The combination of EPC and LTE is referred to as the Evolved Packet System (EPS).

LTE is available in both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes. Many deployments are based on FDD in paired spectrum. The TDD mode, however, is important for deployments in which paired spectrum is unavailable. Instances of TDD deployment include China, Europe at 2.6 GHz, U.S. Broadband Radio Service (BRS) spectrum at 2.6 GHz, and the 3.5 GHz band.

The versions of LTE most widely deployed today are just the first in a series of innovations that will increase performance, efficiency, and capabilities. Enhancements in the 2013 to 2016 period were the ones defined in 3GPP Releases 10, 11, and 12 and are commonly referred to as LTE-Advanced.[20] Subsequent releases, including Releases 13 to 15, specify LTE-Advanced Pro.

Keeping in mind that different operators have varying priorities, the following list roughly ranks the most important features of LTE-Advanced and LTE-Advanced Pro for the 2017 to 2020 timeframe:

1. **Carrier Aggregation.** With this capability, already in use, operators can aggregate radio carriers in the same band or across disparate bands to improve throughputs (under light network load), capacity, and efficiency. Carrier aggregation can also combine FDD and TDD and is the basis of LTE-U and LTE-LAA. As examples, in 2015, AT&T aggregated 700 MHz with AWS, and 700 MHz with PCS. T-Mobile aggregated 700 MHz with AWS, and AWS with PCS.[21] Operators are now deploying three-carrier aggregation and eventually may aggregate four carriers.[22] Release 13 introduced support for carrier aggregation of up to 32 carriers, addressing primarily the opportunity to aggregate multiple unlicensed channels. Release 14 specifies interband carrier aggregation for up to five downlink carriers and 2 uplink carriers.

2. **VoLTE.** Initially launched in 2015 and with widespread availability in 2017, VoLTE enables operators to roll out packetized voice for LTE networks, resulting in greater voice capacity and higher voice quality.

3. **Tighter Integration of LTE with Unlicensed Bands.** LTE-U became available for testing in 2016, and 3GPP completed specifications for LAA in Release 13, with deployment expected around 2018. MulteFire, building on LAA, will operate without requiring a licensed carrier anchor. LTE/Wi-Fi Aggregation through LWA and LWIP are other options for operators with large Wi-Fi deployments.

---

[20] From a strict standards-development point of view, the term "LTE-Advanced" refers to the following features: carrier aggregation, 8X8 downlink MIMO, and 4XN uplink MIMO with N the number of receive antennas in the base station.

[21] AT&T band combinations are 3GPP Band 13 + Band 4, Band 17 + Band 4, and Band 17 + Band 2. T-Mobile band combinations are Band 12 + Band 4, Band 12 + Band 2, and Band 4 + Band 2.

[22] For carrier aggregation to operate, both the network and the device have to support the particular band combination. Legacy devices typically do not support new network aggregation capabilities.

4. **Enhanced Support for IoT.** Release 13 brings Category M1, a low-cost device option, along with Narrowband-IoT (NB-IoT), a version of the LTE radio interface specifically for IoT devices, called Category NB1.

5. **Higher-Order and Full-Dimension MIMO.** Deployments in 2017 use up to 4X4 MIMO. Release 14 specifies a capability called Full-Dimension MIMO, which supports configurations with as many as 32 antennas at the base station. See the section "Smart Antennas and MIMO" and Appendix section "LTE Smart Antennas" for further detail.

6. **Dual Connectivity.** Release 12 introduced the capability to combine carriers from different sectors and/or base stations (i.e. evolved Node Bs [eNBs]) through a feature called Dual Connectivity. Two architectures were defined: one that supports Packet Data Convergence Protocol (PDCP) aggregation between the different eNBs and one that supports separate S1 connections on the user-plane from the different eNBs to the EPC.

7. **256 QAM Downlink and 64 QAM Uplink.** Defined in Release 12 and already deployed in some networks, higher-order modulation increases user throughput rates in favorable radio conditions.

8. **1 Gbps Capability.** Using a combination of 256 QAM modulation, 4X4 MIMO, and aggregation of three carriers, operator networks can now reach 1 Gbps peak speeds. See below for more information.

9. **V2X Communications.** Release 14 specifies vehicle-to-vehicle and vehicle-to-infrastructure communications. See the section "Cellular V2X Communications" for more information.

10. **Coordinated Multi Point.** CoMP (and enhanced CoMP [eCoMP]) is a process by which multiple base stations or cell sectors process a User Equipment (UE) signal simultaneously, or coordinate the transmissions to a UE, improving cell-edge performance and network efficiency. Initial usage will be on the uplink because no user device changes are required. Some networks had implemented this feature in 2017.

11. **HetNet Support.** HetNets integrate macro cells and small cells. A key feature is enhanced inter-cell interference coordination (eICIC), which improves the ability of a macro and a small-cell to use the same spectrum. This approach is valuable when the operator cannot dedicate spectrum to small cells. Operators are currently evaluating eICIC, and at least one operator has deployed it.[23] Further enhanced ICIC (feICIC) introduced in Rel-11 added advanced interference cancellation receivers into devices.

12. **Ultra-Reliable and Low-Latency Communications.** Being specified in Release 15, URLLC in LTE will shorten radio latency to a 1 msec range using a combination of shorter transmission time intervals and faster hybrid automatic repeat request (HARQ) error processing. See the Appendix section "LTE Ultra-Reliable and Low-Latency Communications" for further details.

13. **Self-Organizing Networks.** With SON, networks can automatically configure and optimize themselves, a capability that will be particularly important as small cells begin

---

[23] Fierce Wireless, "SK Telecom teams with Nokia Networks on eICIC," January 2015.

to proliferate. Vendor-specific methods are common for 3G networks, and trials are now occurring for 4G LTE standards-based approaches.

Other key features that will become available in the 2016-to-2020 timeframe include full-dimension MIMO, enhanced Multimedia Broadcast/Multicast Services (eMBMS), User-Plane Congestion Management (UPCON), and device-to-device communication (targeted initially at public-safety applications).

The appendix explains these features and quantifies performance gains, and Figure 10 illustrates the transition from LTE to LTE-Advanced and LTE-Advanced Pro, which include these features.

Figure 10: LTE to LTE-Advanced Pro Migration[24]
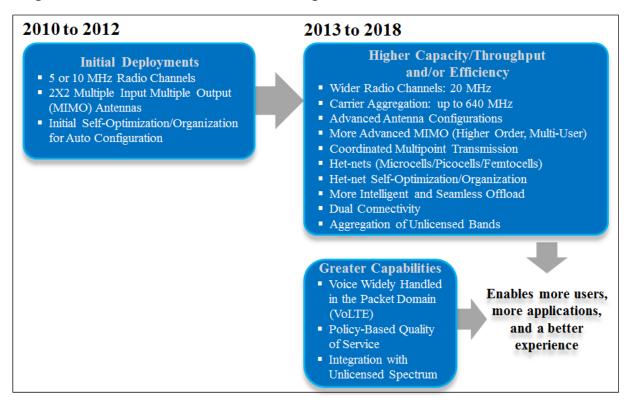
Figure 11 illustrates gains in peak downlink speeds through carrier aggregation, higher-order MIMO, and higher-order modulation.

---

[24] 5G Americas/Rysavy Research
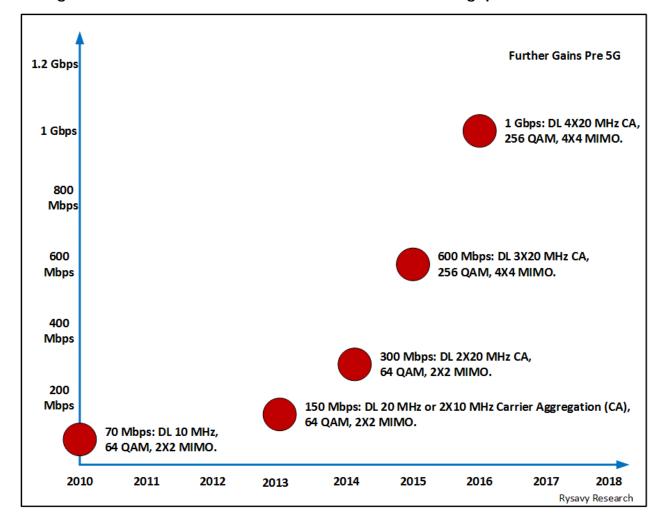
Figure 11: Successive Gains in Peak LTE Downlink Throughput[25]

5G Use Cases (ITU and 3GPP)
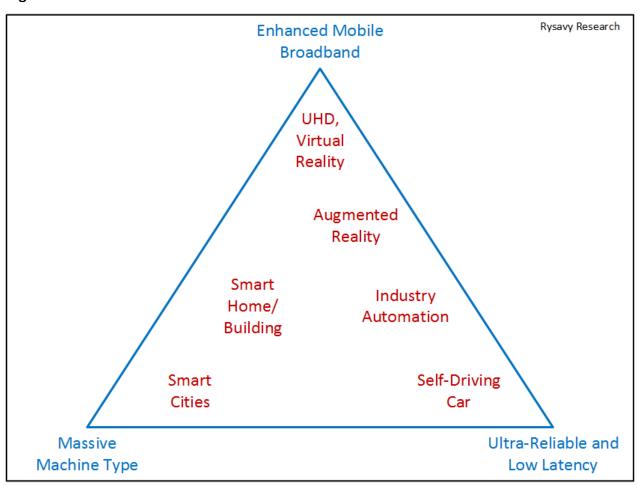
The ITU, in its 5G recommendations, divides use cases into three main categories, as shown in Figure 12.

1. **Enhanced Mobile Broadband (eMBB).** eMBB is the most obvious extension of LTE capability, providing higher speeds for applications such as streaming, Web access, video conferencing, and virtual reality. Highest speeds will occur in small cells with limited movement speed of end users, such as with pedestrians.

2. **Massive Machine-Type Communications (mMTC).** Massive machine-type communications extends LTE Internet of Things capabilities—for example, NB-IoT—

---

[25] Rysavy Research analysis based on information about commercially available chipsets. Not all LTE features are available simultaneously, and thus throughputs do not necessarily scale linearly with items such as bandwidth. See also the following article about achieving gigabit speeds with LTE: Ars Technica, "Qualcomm's new LTE modem will make gigabit download speeds easier to hit," February 21, 2017. Available at https://arstechnica.com/gadgets/2017/02/qualcomms-new-lte-modem-will-make-gigabit-download-speeds-easier-to-hit/#p3.

to support huge numbers of devices with lower costs, enhanced coverage, and long battery life. As shown in the ITU objectives, below, 5G will support ten times as many devices in an area as LTE.

3. **Ultra-Reliable and Low Latency Communications (URLLC).** Of the three categories, URLLC enables wireless applications never before possible. Driven by high dependability and extremely short network traversal time, URLLC, also referred to as "mission-critical" communications, will enable industrial automation, drone control, new medical applications, and autonomous vehicles. These types of applications are potentially the ones that will deliver the greatest societal benefits, yet unfortunately, at least in the United States, they could possibly be undermined by current network neutrality regulations. This category is also referred to as critical machine-type communications (cMTC).

Figure 12: ITU Use Case Model[26]

---

[26] For background, see ITU, *IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond*, Recommendation ITU-R M.2083-0, Sep. 2015.

3GPP, in studying 5G, has methodically identified multiple specific use cases in a project called "SMARTER," which are consistent with ITU's model.[27] 3GPP's service dimensions include:

❑ Massive Internet of Things (eHealth, wearables, eCity, eFarm).

❑ Critical Communications (e.g., vehicles, drones, industry robots).

❑ Enhanced Mobile Broadband (e.g., augmented reality, virtual reality, ultra-high definition).

❑ Network Operation (e.g., network slicing, connectivity/routing, migration/interworking).

Using small cells and mmWave radio channels, a 5G network built for capacity will deliver 1 Tbps/km^2 or higher, enabling 5G to compete with wireline broadband services.[28]

5G Technical Objectives

Table 4 shows the ITU's objectives for IMT-2020 (5G) relative to IMT-Advanced (4G).

Table 4: ITU Objectives for IMT-2020 compared with IMT-Advanced[29]

	IMT-Advanced	IMT-2020
Peak Data Rate	DL: 1 Gbps UL: 0.05 Gbps	DL: 20 Gbps UL: 10 Gbps
User Experienced Data Rate	10 Mbps	100 Mbps[30]
Peak Spectral Efficiency	DL: 15 bps/Hz UL: 6.75 bps/Hz	DL: 30 bps/Hz UL: 15 bps/Hz
Average Spectral Efficiency		DL eMBB indoor: 9 bps/Hz DL eMBB urban: 7.8 bps/Hz DL eMBB rural: 3.3 bps/Hz UL eMBB indoor: 6.75 bps/Hz UL eMBB urban: 5.4 bps/Hz UL eMBB rural: 1.6 bps/Hz

---

[27] 3GPP TR22.891, *Feasibility Study on New Services and Markets Technology Enablers*; TR22.861 (Massive Internet of Things); TR22.862 (Critical Communications); TR 22.863 (Enhanced Mobile Broadband); TR22.864 (Network Operation).

[28] The hotspot capacity requirement of 10 Mbps/sq. m. is equivalent to 10 Tbps/sq. km. See also *Nokia, Ten key rules of 5G deployment, Enabling 1 Tbit/s/km^2 in 2030*, 2015.

[29] ITU Working Party 5D, *Minimal Requirements Related to Technical Performance for IMT-2020 Radio Interfaces*, Feb 22, 2017. See also 3GPP TR 38.913, *Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14)*, V14.2.0, Mar. 2017.

[30] Per ITU, "User experienced data rate is the 5% point of the cumulative distribution function (CDF) of the user throughput."

	IMT-Advanced	IMT-2020
Mobility	350 km/h	500 km/h
User Plane Latency	10 msec	1 msec[31]
Connection Density	100 thousand devices/sq.km.	1 million devices sq./km.
Network Energy Efficiency	1 (normalized)	100X over IMT-Advanced
Area Traffic Capacity	0.1 Mbps/sq. m.	10 Mbps/sq. m. (hot spots)
Bandwidth	Up to 20 MHz/radio channel (up to 100 MHz aggregated)	Up to 1 GHz (single or multipole RF carriers)

In supporting different usage scenarios, not all of these objectives will necessarily be simultaneously available. For example, an IoT application may need to support a large number of devices but at lower throughput rates, while a vehicular application may need high mobility and low latency.

Other expected enhancements include:

❑ Deep coverage for machines buried within environments.

❑ Extremely low energy demands for ten years or more of battery operation.

❑ Low complexity options for inexpensive machine communications.

❑ Auto-awareness through discovery and self-optimization.

5G Concepts and Architectures

Standards bodies have not yet defined 5G requirements, but various groups are analyzing the possibilities of what might constitute 5G for network deployments in 2020 or beyond. Often stated goals of 5G include:
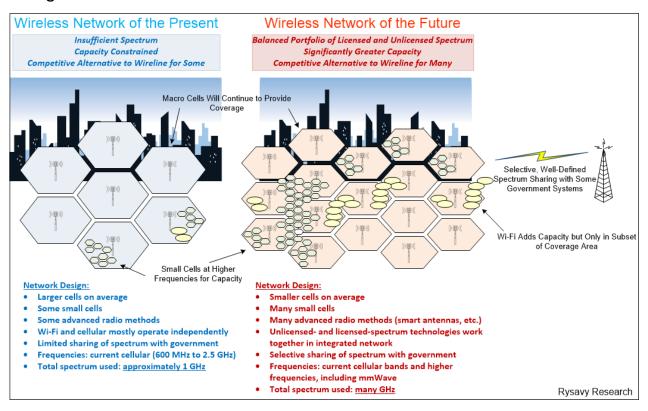
❑ Being able to support a greater number of end-systems, including IoT applications, at lower average revenue than 4G systems.

❑ Peak data rates of multi Gbps (see Table 4 above).

❑ Rather than emphasizing peak rates, a more uniform user experience across the coverage area.

❑ Support for many frequencies, including existing cellular bands and frequencies above 6 GHz.

---

[31] Per 3GPP TR 38.913 (V14.2.0, Mar. 2017), 0.5 msec for DL and 0.5 msec for UL for URLCC and 4 msec for UL and 4 msec for DL for eMBB.

❑ Availability of TDD and FDD modes for all bands.

❑ Hierarchical/planned and ad hoc deployment models.

❑ Use of licensed and unlicensed bands.

❑ Equal support for human-type and machine-type communications. Includes highly efficient small data transmission.

❑ Advanced Spectrum Sharing, possibly based on spectrum sharing approaches being developed for the 3.5 GHz band.

Figure 13 shows the transformation of networks, moving from today's LTE-Advanced networks to future LTE-Advanced and eventually 5G networks.

Figure 13: Network Transformation[32]
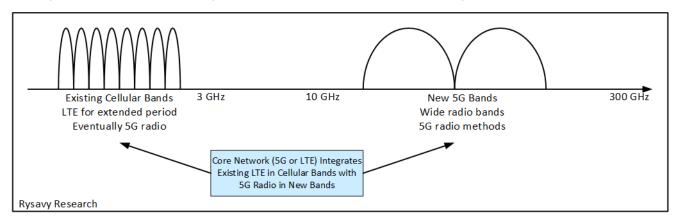
The fundamental decision for 5G is how to best leverage existing technology investments in LTE while exploiting new spectrum and new technology capabilities. 5G design emphasizes ways to combine existing 4G LTE networks with capabilities provided by 5G. One approach likely to be used by many operators is to use LTE in existing frequency bands and the 5G NR in new bands, such as mmWave, as shown in Figure 14. 5G NR, however,

---

[32] See also Rysavy Research infographic, "Mobile Broadband Networks of the Future," April 2014. Available at http://www.rysavy.com/Articles/2014-05-Networks-of-the-Future-Infographic.pdf.
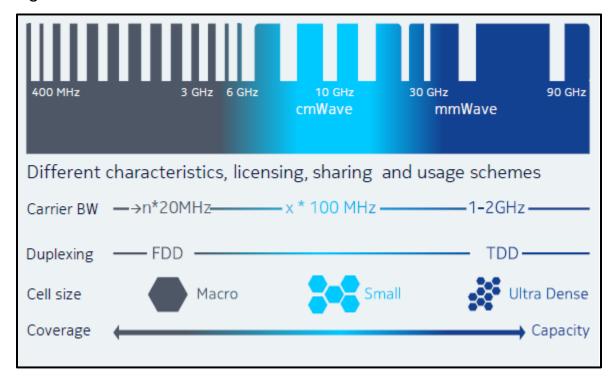
will operate in all frequencies, and just as 2G and 3G spectrum has been re-farmed for LTE, existing cellular bands will eventually be re-farmed for 5G.

Figure 14: 5G Combining of LTE and New Radio Technologies

As shown in Figure 15, higher frequency bands in 5G will provide capacity with smaller cells, and lower bands will provide coverage with larger cells. This is similar to the approach taken in 4G.

Figure 15: Characteristics of Different Bands[33]

---

[33] Nokia, *Vision & Priorities for Next Generation Radio Technology*, 3GPP RAN workshop on 5G, Sep. 17-18, 2015.

New radio methods in NR will boost performance and will co-exist with future versions of LTE. 3GPP, in Release 14, completed a study item on 5G NR and will specify NR in Release 15.

Figure 16: Evolution to 5G Including LTE Improvements and New 5G Radio Methods

While many deployments will integrate LTE and NR, some operators could choose NR-only deployments.

5G New Radio (NR)

3GPP is defining 5G NR in Release 15, with an initial release for non-standalone operation scheduled for March 2018 and full release for standalone version scheduled for September 2018. Release 16 will add additional capabilities to the radio, such as support for URLLC. Despite specifications not being finalized, the following are planned capabilities and features of NR in Release 15:

❑ Ability to operate in any frequency band, including low, mid, and high bands.

❑ Network can support both LTE and 5G NR, including dual connectivity with which devices have simultaneous connections to LTE and NR.

❑ Carrier aggregation for multiple NR carriers.

❑ 5 Gbps peak downlink throughput in initial releases, increasing to 50 Gbps in subsequent versions.

❑ OFDMA in downlink and uplink, with optional Single Carrier Frequency Division Multiple Access (SC-FDMA) for uplink. Radio approach for URLLC to be defined in

Release 16, but Release 15 will provide physical layer frame structure and numerology support.

❑ Massive MIMO and beamforming.

❑ Ability to support either FDD or TDD modes for 5G radio bands.

❑ Numerologies of 2^N X 15 kHz for subcarrier spacing up to 120 kHz or 240 kHz. This approach, depicted in Figure 17, supports both narrow radio channels (for example, 1 MHz), or wide ones (for example, 400 MHz). Phase 1 likely to support a maximum of 400 MHz bandwidth with 240 kHz subcarrier spacing.

❑ Error correction through low-density parity codes (LDPC), which are computationally more efficient than LTE turbo codes at higher data rates.

❑ Standards-based cloud RAN support, compared with proprietary LTE approaches, that specifies a split between the PDCP and Radio Link Control (RLC) protocol layers.

❑ Self-contained integrated subframes that combine scheduling, data, and acknowledgement.

❑ Future-proofing by providing a flexible radio framework that has forward compatibility to support future, currently unknown services, such as URLLC to be specified in Release 16.

❑ Scalable transmission time intervals with short time intervals for low latency and longer time intervals for higher spectral efficiency.

❑ QoS support using a new model.

❑ Dynamic co-existence with LTE in the same radio channels. (See the Appendix section "LTE-NR Co-existence" for more details.)

According to one 5G Americas member company in 2017, operators globally have expressed interest in deploying NR in a wide variety of bands, including current cellular bands, 3.5 GHz, and mmWave bands.

Release 16 will add support for:

❑ URLCC.

❑ Shared spectrum operation.

❑ 5 GHz unlicensed spectrum operation based on current LTE approaches such as LAA, LWA, LWIP, and RAN Controlled LTE Wireless LAN Interworking (RCLWI).

❑ Possibly integrated access and backhaul based on a Release 15 study item. See the appendix section, "Integrated Access and Backhaul," for more details.

❑ Other, as yet unknown, features.

The ability to simultaneously transmit and receive on the same frequency has been stated in the past as an objective of 5G, and although such capability remains of interest, it is not currently being specified.
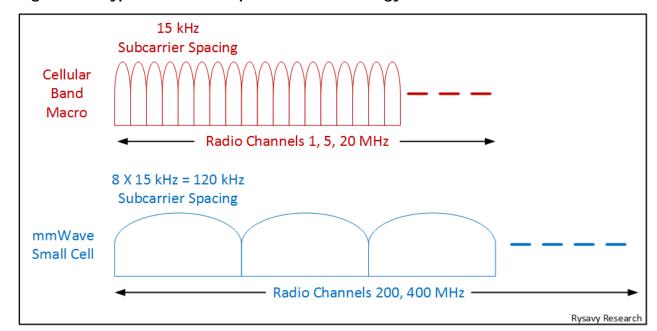
Figure 17: Hypothetical Example of 5G Numerology

Expected 5G Performance

Predicting the 5G user experience is challenging because 5G will be deployed in many configurations, including different bands and with varying width radio channels. In addition, the throughput rates a user experiences depend on signal quality, device capability, and network loading. Some early predictions, however, are possible by examining the ITU objectives, the expected width of radio channels, and results from simulation studies.

The ITU objectives are that in a 5G network, 95% or more users should be able to experience at least 100 Mbps, which is also termed as cell-edge throughput. Examining the simulation results in the appendix section on 5G, "Performance," such results appear feasible using 200 MHz or wider radio channels. The results also suggest that average or typical rates could reach 500 Mbps using 200 MHz radio channels and 1 Gbps using 400 MHz radio channels.

Just as LTE throughputs have increased significantly over this decade, 5G performance will keep improving over the next ten years. One 5G Americas member expects that the 5G theoretical peak rate could be 5 Gbps in 2020, rising to 50 Gbps by 2024.

Release 15 Non-Standalone and Standalone Options

3GPP is specifying the first phase of 5G in Release 15. So that operators can deploy 5G sooner, 3GPP divided Release 15 into two sets of specifications. The first and earlier set of specifications define how a 5G RAN can integrate with an LTE network in what 3GPP calls a non-standalone option. In this earliest version (architecture option 3), NR will rely on an existing LTE network, both in the RAN and in the core.

The complete Release 15 specifications will also define a 5G core network. Figure 18 shows some of the architecture options. Options 3, 4 and 7 are the non-standalone options and options 1, 2, and 5 are standalone. The appendix section, "5G Architecture Options," discusses all of the deployment options in greater detail.
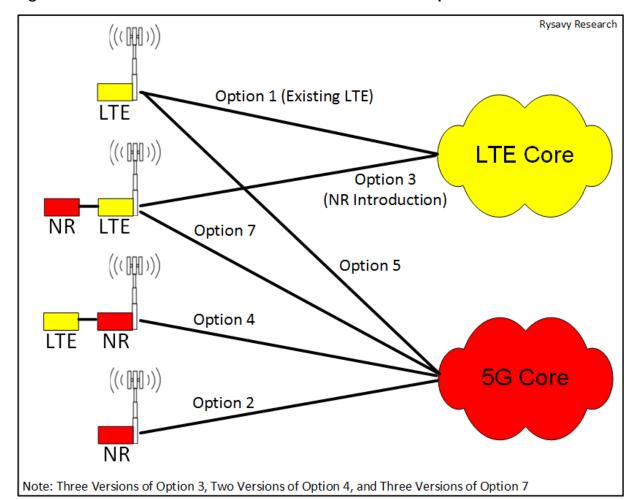
Figure 18: **Release 15 Non-Standalone and Standalone Options**

Note: Three Versions of Option 3, Two Versions of Option 4, and Three Versions of Option 7

5G Schedule3GPP is currently standardizing 5G in Release 15, which will complete the non-standalone version of 5G in March 2018. Based on a typical minimum period of 18 months to build and deploy the technology, initial 5G NSA deployments could occur toward the end of 2019 or the beginning of 2020. 3GPP will complete the full Release 15 specifications in September 2018, enabling deployments in 2020. Release 16, which is the second phase of 5G, will be complete at the end of 2019, and Release 16 deployments could occur in 2021. In 2020, 3GPP will begin work on Release 17 which will include as yet unknown capabilities.

Figure 19Figure 19 shows the current schedule for 5G development and deployment. 3GPP is currently standardizing 5G in Release 15, which will complete the non-standalone version

of 5G in March 2018. Based on a typical minimum period of 18 months to build and deploy the technology, initial 5G NSA deployments could occur toward the end of 2019 or the beginning of 2020. 3GPP will complete the full Release 15 specifications in September 2018, enabling deployments in 2020. Release 16, which is the second phase of 5G, will be complete at the end of 2019, and Release 16 deployments could occur in 2021. In 2020, 3GPP will begin work on Release 17 which will include as yet unknown capabilities.
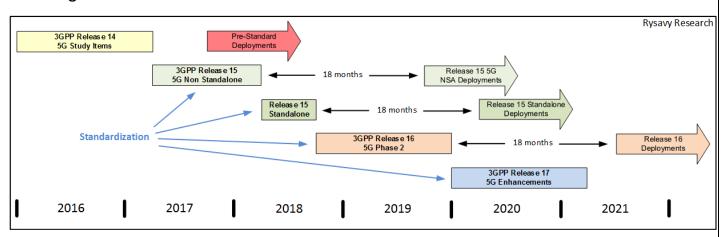
Figure 19: 5G Schedule[34]

5G Core Network

Release 15 also defines initial core network capabilities that support QoS and network slicing. See the 5G appendix section, "Architecture," for more details. A study item in Release 15 will study integrated access and backhaul, with which base stations can use the 5G radio channel for backhaul. This capability could become a work item for Release 16. See the appendix section, "Integrated Access and Backhaul," for more details.

Many operators will virtualize their 5G core networks, just as they have for LTE, but such virtualization is outside the scope of 3GPP specifications.

5G Device Availability

User devices capable of 5G operation have not yet been announced, but availability will likely follow the trends of previous generations of networks. Initial devices, possibly in the 2019 timeframe, will likely include routers that have a 5G radio and use Wi-Fi for local Hotspot capability and USB modems. Handset vendors are in the early stages of designing mmWave support into smartphones.[35] These devices could come online in the 2021 timeframe, although this estimate could tighten or lengthen depending on chipset availability and handset vendor plans.

---

[34] 3GPP and 5G Americas member contribution. Completion based on anticipated ASN.1 freeze dates.

[35] For example, see "Wireless Week, Apple Wants to Test Performance at 28 GHz, 39 GHz for 5G Devices," May 24, 2017, https://www.wirelessweek.com/news/2017/05/apple-wants-test-performance-28-ghz-39-ghz-5g-devices, viewed May 29, 2017.

mmWave

Use of higher frequencies, such as above 6 GHz, represents one of the greatest opportunities for higher throughputs and higher capacity. This benefit derives from the potential availability of ten times the amount of spectrum as is currently available, with multiple GHz of contiguous spectrum. But these higher frequencies, especially mmWave frequencies (above 30 GHz), are suitable only over short distances, as explained below. The combination of lower and higher frequencies is therefore crucial for 5G operation. Lower bands can be devoted to coverage and control, while higher bands can provide opportunistic access for high data rates. The lower and higher spectrum bands can operate in a carrier aggregation or dual connectivity model. Initial 5G specifications include such dual-connectivity capability.

Compared with lower frequencies, mmWave frequencies suffer from worse propagation characteristics, even in line-of-sight conditions, because the comparatively smaller aperture area of the receiver's antenna requires some form of beamforming at the transmit side, and potentially even at the receive side. Fortunately, the smaller form factors of mmWave antennas allow for dense packing of antenna arrays. Experimental systems using antenna arrays have demonstrated reliable communications at 28 GHz, even in dense, urban, non-line-of-sight conditions, for distances up to 200 meters.[36] Arrays at the terminal side are space-constrained, but some basic beamforming at the terminal is possible. On the base station side, the arrays may include hundreds of antennas in an approach called "massive MIMO."

Figure 20 shows how an increasing number of antenna elements can extend coverage through tighter beams. A 77 X 77 antenna array (6,000 elements) can exceed a kilometer at 3.5 GHz (33 dBm transmit power) and reach over 800 meters, even at 30 GHz.

---

[36] Samsung, *5G Vision,* February 2015. Available at http://www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-0.pdf.
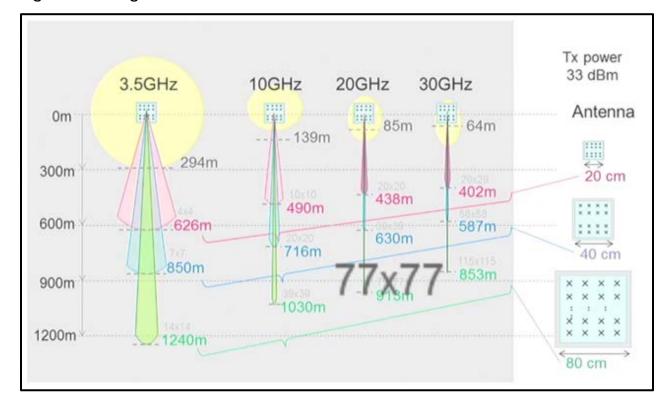
Figure 20: Range Relative to Number of Antenna Elements[37]

More typically, mmWave cells will employ shorter ranges of 50 to 200 meters. Extreme densification is another way that 5G networks will achieve massive increases in capacity. 3G networks reached densities of four to five base stations per sq. km, 4G networks eight to 10, but 5G networks may reach densities of 40 to 50.[38] A likely 5G architecture will use the macro cell for control information, coverage, and fallback, but small cells, often operating at higher frequencies, for high-bandwidth data communication. Either wireless connections or fiber will provide backhaul. Figure 21 shows how such an approach could also employ beamforming and beam tracking when using mmWave bands in the small cells.

---

[37] Dr. Seizo Onoe, NTT DOCOMO, presentation at Brooklyn 5G Summit, Apr. 21, 2016. Used by permission.

[38] IEEE Wireless Communications, *5G Ultra-Dense Cellular Networks,* Feb. 2016.
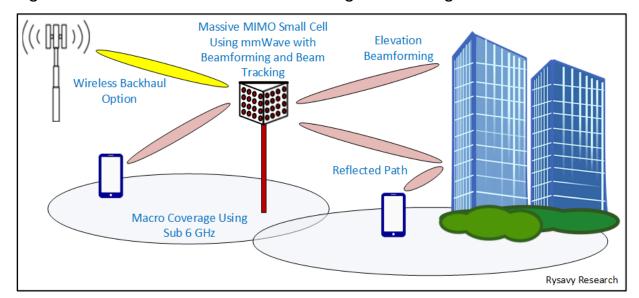
Figure 21: 5G Architecture for Low-Band/High-Band Integration

In combination, the various methods expected in 5G will provide users in mmWave band hotspot coverage at least a 100-fold increase in throughput over LTE, achieved by:

- ❑ Five- to 10-fold gains due to fewer users in each small cell. (Five to ten times as many cells.)

- ❑ 10-fold gains from access to much larger amounts of spectrum.

- ❑ Three-fold gains or more from improved spectral efficiency.

It is this huge increase in capacity, combined with Gbps performance, that will allow 5G to compete with wireline networks.[39]

Network Slicing

Not only will 5G networks include a new radio, but thanks to virtualization, these networks will be able to present multiple faces for different use cases using a concept called network slicing. This architecture allows an operator to provide multiple services with different performance characteristics. Each network slice operates as an independent, virtualized version of the network. For an application, the network slice is the only network it sees. The other slices, to which the customer is not subscribed, are invisible and inaccessible. The advantage of this architecture is that the operator can create slices that are fine-tuned for specific use cases. One slice could target autonomous vehicles, another enhanced mobile broadband, another low-throughput IoT sensors, and so on.

Figure 22 shows the network slicing architecture, with devices having access to only the slice or slices for which they have subscriptions. Each slice has radio resources allocated,

---

[39] For a further discussion of 5G capacity and ability to compete with wireline networks, refer to Datacomm Research and Rysavy Research, *Broadband Disruption: How 5G Will Reshape the Competitive Landscape*, 2017, available at https://datacommresearch.com/reports-broadband/.

with specific QoS characteristics. Within the core network, virtualized core network functions support each slice and provide connections to external networks. [40]
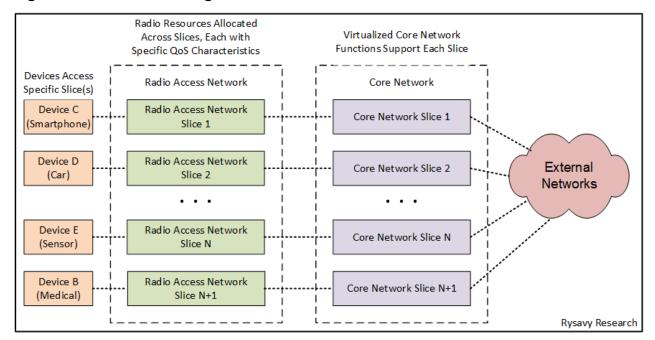
Figure 22: Network Slicing Architecture

Information-Centric Networking

For many usage scenarios, wireless networks provide broadband access to the internet, a network that itself is evolving. The internet is based on a node-centric design developed forty years ago. The point-to-point method of communication the internet uses has functioned well for a vast array of applications but is not optimal for the way content is developed and distributed today. Industry and academic organizations are researching a concept called "Information-Centric Networking." ICN seeks a new approach of in-network caching that distributes content on a large scale, cost-efficiently and securely.

Most internet content uses Uniform Resource Identifiers (URIs) to locate objects and define specific location-dependent IP addresses. This approach, however, causes problems when content moves, sites change domains, or content is replicated and each copy appears as a different object. Developments such as peer-to-peer overlays and content distribution networks (such as Akamai) that distribute cached copies of content are a first step toward an information-centric communication model.

ICN is built from the ground up on the assumption of mobility, so it eliminates the mobility overlays on which current mobile broadband networks depend. The approach will be able to place information anywhere in the network with immediate and easy retrieval.

Key principles of ICN include:

---

[40] For more details, see 5G Americas, *Network Slicing for 5G Networks & Services*, November 2016. Available at:
http://www.5gamericas.org/files/3214/7975/0104/5G_Americas_Network_Slicing_11.21_Final.pdf.

- ❑ The architecture inherently supports user mobility.

- ❑ Network operations are name-based instead of address- or node-based.

- ❑ The network itself stores, processes, and forwards information.

- ❑ Intrinsic security guarantees the integrity of every data object.

The goal of ICN is to simplify the storage and distribution of gigantic amounts of content while reducing the amount of traffic and latency users face when accessing the content. The internet cannot just be replaced, however, so in initial stages, ICN would operate as an overlay, and over time would assume an increasing percentage of the functions within the internet. ICN would not discard IP; rather, it seeks to generalize the routing concept to enrich networking with new capabilities.

Some technology aspects of ICN include:

- ❑ Information retrieval from multiple sources without needing to know the location of the information.

- ❑ Multipath communications that improves user performance and traffic load balancing.

- ❑ Subsequent requests for the same data will be served locally without needing to fetch it from original repository.

- ❑ Elimination of the name-to-location indirection associated with Domain Name Service (DNS).

Because mobility is such a central aspect of ICN, mobile network operators are in a unique position to participate in ICN-related research and development, and to do so as part of 5G development. ICN has not progressed to a level at which 3GPP specification work could include it, so instead promoters are ensuring that 5G specification work does not preclude it. With this approach, operators in the 2020s will have the option of overlaying ICN capability on their 5G networks. ICN could even be implemented as a 5G network slice for mobile and end-systems capable of ICN.

3GPP Releases

3GPP standards development falls into three principal areas: radio interfaces, core networks, and services. Progress in the 3GPP family of technologies has occurred in multiple phases, first with GSM, then GPRS, EDGE, UMTS, HSPA, HSPA+, LTE, LTE-Advanced, and now LTE-Advanced Pro. Underlying radio approaches have evolved from Time Division Multiple Access (TDMA) to CDMA to Orthogonal Frequency Division Multiple Access (OFDMA), which is the basis of LTE. 3GPP is also involved in standardization of 5G technology.

Table 5 summarizes the key 3GPP technologies and their characteristics.

Table 5: Characteristics of 3GPP Technologies

Technology Name	Type	Characteristics	Typical Downlink Speed	Typical Uplink Speed
HSPA[41]	WCDMA	Data service for UMTS networks. An enhancement to original UMTS data service.	1 Mbps to 4 Mbps	500 Kbps to 2 Mbps
HSPA+	WCDMA	Evolution of HSPA in various stages to increase throughput and capacity and to lower latency.	1.9 Mbps to 8.8 Mbps in 5+5 MHz[42] 3.8 Mbps to 17.6 Mbps with dual-carrier in 10+5 MHz	1 Mbps to 4 Mbps in 5+5 MHz or in 10+5 MHz
LTE	OFDMA	New radio interface that can use wide radio channels and deliver extremely high throughput rates. All communications handled in IP domain.	6.5 to 26.3 Mbps in 10+10 MHz[43]	6.0 to 13.0 Mbps in 10+10 MHz
LTE-Advanced	OFDMA	Advanced version of LTE designed to meet IMT-Advanced requirements.	Significant gains through carrier aggregation, 4X2 and 4X4 MIMO, and 256 QAM modulation.	

User-achievable rates and additional details on typical rates are covered in the appendix section "Data Throughput."

3GPP develops specifications in releases, with each release addressing multiple technologies. For example, Release 8 defined dual-carrier operation for HSPA but also introduced LTE. Each release adds new features and improves performance of existing functionality in different ways. Table 6 summarizes some key features of different 3GPP releases.

---

[41] HSPA and HSPA+ throughput rates are for a 5+5 MHz deployment.

[42] "5+5 MHz" means 5 MHz used for the downlink and 5 MHz used for the uplink.

[43] 5G Americas member company analysis for downlink and uplink. Assumes single user with 50% load in other sectors. AT&T and Verizon are quoting typical user rates of 5-12 Mbps on the downlink and 2-5 Mbps on the uplink for their networks. See additional LTE throughput information in the section below, "LTE Throughput."

Table 6: Key Features in 3GPP Releases[44]

Release	Year	Key Features
99	1999	First deployable version of UMTS.
5	2002	High Speed Downlink Packet Access (HSDPA) for UMTS.
6	2005	High Speed Uplink Packet Access (HSUPA) for UMTS.
7	2008	HSPA+ with higher-order modulation and MIMO.
8	2009	Long Term Evolution. Dual-carrier HSDPA.
10	2011	LTE-Advanced, including carrier aggregation and eICIC.
11	2013	Coordinated Multi Point (CoMP).
12	2015	Public safety support. Device-to-device communications. Dual Connectivity. 256 QAM on the downlink.
13	2016	LTE-Advanced Pro features. LTE operation in unlicensed bands using LAA. Full-dimension MIMO. LTE-WLAN Aggregation. Narrowband Internet of Things.
14	2017	LTE-Advanced Pro additional features, such as eLAA (adding uplink to LAA) and cellular V2X communications. Study item for 5G "New Radio."
15	2018	Additional LTE-Advanced Pro features, such as ultra-reliable low-latency communications. Phase 1 of 5G. Will emphasize enhanced mobile broadband use case and sub-40 GHz operation. Includes Massive MIMO, beamforming, and 4G-5G interworking, including ability for LTE connectivity to a 5G CN.
16	2019	Phase 2 of 5G. Full compliance with ITU IMT-2020 requirements. Will add URLLC, spectrum sharing, unlicensed operation, and multiple other enhancements.
17	2021	Further LTE and 5G enhancements.

Refer to the Appendix section "3GPP Releases" for a more detailed listing of features in each 3GPP Release.

---

[44] After Release 99, release versions went to a numerical designation beginning with Release 4, instead of designation by year.

Supporting Technologies and Architectures

Network architects design networks using a deep and wide toolkit, including multiples types of cell sizes, integration with unlicensed spectrum, smart antennas, converged services, and virtualization.

Multiple Cell Types

Operators have many choices for providing coverage. Lower frequencies propagate further and thus require fewer cells for coverage. The resulting network, however, has lower capacity than one with more cells, so operators must continually evaluate cell placement with respect to both coverage and capacity.

Table 7 lists the many types of cells. Note that the distinctions, such as radius, are not absolute—perhaps one reason the term "small cell" has become popular, as it encompasses picocells, metrocells, femtocells, and sometimes Wi-Fi.

With "plug-and-play" capability derived from self-configuring and self-organizing features, small cells will increasingly be deployed in an ad hoc manner, anywhere power and backhaul are available, yet will operate in tight coordination with the rest of the network.

A proliferation of small cells inside buildings will also provide coverage from inside to outside, such as in city streets, the reverse of traditional coverage that extends from outdoor cells to inside.

Table 7: Types of Cells and Typical Characteristics (Not Formally Defined)

Type of Cell	Characteristics
Macro cell	Wide-area coverage. LTE supports cells up to 100 km in range, but typical distances are .5 to 5 km radius. Always installed outdoors.
Microcell	Covers a smaller area, such as a hotel or mall. Range to 2 km, 5-10W, and 256-512 users. Usually installed outdoors.
Picocell	Indoor or outdoor. Outdoor cells, also called "metrocells." Typical range 15 to 200 meters outdoors and 10 to 25 meters indoors, 1-2W, 64-128 users. Deployed by operators primarily to expand capacity.
Consumer Femtocell	Indoors. Range to 10 meters, less than 50 mW, and 4 to 6 users. Capacity and coverage benefit. Usually deployed by end users using their own backhaul.
Enterprise Femtocell	Indoors. Range to 25 meters, 100-250 mW, 16-32 users. Capacity and coverage benefit. Deployed by operators.
Distributed antenna system	Expands indoor or outdoor coverage. Same hardware can support multiple operators (neutral-host) since antenna can support broad frequency range and multiple technologies. Indoor deployments are typically in larger spaces such as airports. Has also been deployed outdoors for coverage and capacity expansion.

Type of Cell	Characteristics
Remote radio head (RRH)	Uses baseband at existing macro site or centralized baseband equipment. If centralized, the system is called "cloud RAN." Requires fiber connection.
Wi-Fi	Primarily provides capacity expansion. Neutral-host capability allows multiple operators to share infrastructure.

Smalls Cells and Heterogeneous Networks

Historically, increasing the number of cell sites has been the primary method for increasing capacity, providing gains far greater than what can be achieved by improvements in spectral efficiency alone. The next wave of densification is by using what the industry calls "small cells."

Central to small-cell support is the heterogeneous network architecture, with multiple types of cells serving a coverage area, varying in frequencies used, radius, and even radio technology used.

HetNets offer significant increases in capacity and improvements, including:

1. Smaller cells, such as open femtocells (home-area coverage) and picocells (city-block-area coverage), inherently increase capacity because each cell serves a smaller number of users.

2. Strategic placement of picocells within the macro cell provides the means to absorb traffic in areas where there are higher concentrations of users. Locations can include businesses, airports, stadiums, convention centers, hotels, hospitals, shopping malls, high-rise residential complexes, and college campuses.

3. Smaller cells can also improve signal quality in areas where the signal from the macro cell is weak.
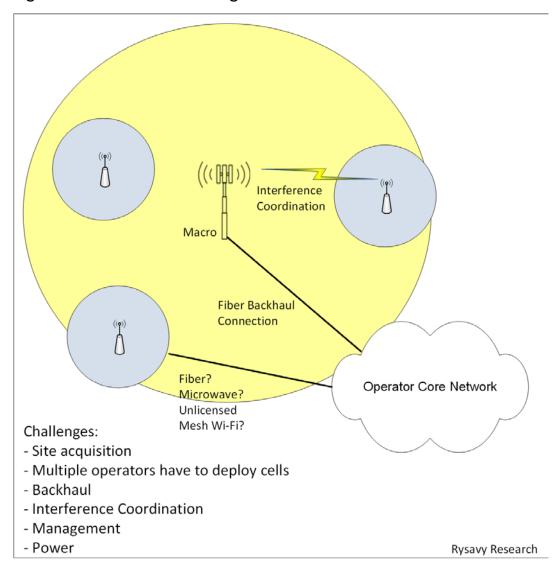
Essential elements for practical HetNet deployment are self-optimization and self-configuration, especially as the industry transitions from tens of thousands of cells to hundreds of thousands, and eventually to millions. The appendix covers technical aspects of HetNets in the sections, "Heterogeneous Networks and Small Cells" and "Self-Organizing Networks."

While promising in the long term, one immediate challenge in deploying a large number of small cells is backhaul, since access to fiber is not necessarily available and line-of-sight microwave links are not always feasible. Site acquisition and the need for multiple operators to deploy their own cells in a coverage area are additional challenges.[45] Figure 23 depicts the challenges.

---

[45] For further discussion of this topic, refer to 5G Americas and Small Cell Forum, *Small cell siting challenges,"* February 2017.

Figure 23: Small-Cell Challenges

Interference Coordination

Macro

Fiber Backhaul Connection

Operator Core Network

Fiber?
Microwave?
Unlicensed
Mesh Wi-Fi?

Challenges:
- Site acquisition
- Multiple operators have to deploy cells
- Backhaul
- Interference Coordination
- Management
- Power

Rysavy Research

Despite these challenges and the relatively small number of small cells deployed today, many believe small-cell deployments will accelerate.[46]

In the United States, the FCC in April 2017 issued a Notice of Proposed Rulemaking, titled "Accelerating Wireless Broadband Deployment by Removing Barriers to Infrastructure

---

[46] For example, see Fierce Wireless, "Crown Castle expects its small cell business to double in next 2 years," April 25, 2017. Available at http://www.fiercewireless.com/wireless/crown-castle-expects-its-small-cell-business-to-double-next-two-years.

Investment," that seeks to streamline items such as site acquisition and consistent pricing.[47]

5G small-cell considerations include:

❑ Due to limited propagation at mmWave frequencies, 5G small-cell deployments will be dense and involve large numbers of sites. Inter-site distances (ISDs) will range from 100 to 300 meters in many deployments, with 200 meters a typical value.[48]

❑ The high-capacity of mmWave small cells will require multi-Gbps backhaul connections using an expected combination of fiber, mmWave radio in point-to-point connections, and 5G self-backhaul.

❑ The expected use of cloud RAN and centralized base station facilities will simplify equipment at the site, facilitating dense deployments.

❑ Dense deployments will motivate neutral-host (multi-tenant) approaches, but these are outside the scope of specification efforts.

The effective range of a mmWave small cell depends on multiple factors, including whether line-of-sight is available, extent of foliage, pole height, whether user equipment is indoors or outdoors, and the types of building materials the signal must pass through to reach indoor equipment.

Despite the challenges, small cells will ultimately contribute greatly to increased network capacity. Table 8 lists possible configurations. Note that many of these approaches can be combined, such as using picos and Wi-Fi offload.

Table 8: Small-Cell Approaches

Small-Cell Approach	Characteristics
Macro plus small cells in select areas.	Significant standards support. Femtocells or picocells can use the same radio carriers as macro (less total spectrum needed) or can use different radio carriers (greater total capacity).
Macro in licensed band plus LTE operation in unlicensed bands.	Promising approach for augmenting LTE capacity in scenarios where operator is deploying LTE small cells.[49] See discussion below in the section on unlicensed spectrum integration.

---

[47] FCC, *Wireless Infrastructure NPRM and NOI*, April 21, 2017. Available at https://www.fcc.gov/document/wireless-infrastructure-nprm-and-noi.

[48] 5G Americas member contributions.

[49] See Rysavy Research, *Accelerating Innovation in Unlicensed Spectrum,* Fierce Wireless, November 2016. Available at http://www.rysavy.com/Articles/2016-11-Innovation-Unlicensed-Spectrum.pdf. See also Rysavy Research, *Will LTE in Unlicensed Spectrum Unlock a Vast Store of Mobile Broadband Capacity?* MIMO World, June 2014. Available at http://www.mimoworld.com/?p=2377.

Small-Cell Approach	Characteristics
Macro (or small-cell) cellular in licensed band plus Wi-Fi.	Extensively used today with increased use anticipated. Particularly attractive for expanding capacity in coverage areas where Wi-Fi infrastructure exists but small cells with LTE do not. LTE Wi-Fi Aggregation (being specified in Release 13) is another approach, as are MP-TCP and MP-QUIC.
Wi-Fi only.	Low-cost approach for high-capacity mobile broadband coverage, but impossible to provide large-area continuous coverage without cellular component.

Neutral-Host Small Cells

Multi-operator and neutral-host solutions could accelerate deployment of small cells.[50] Currently, nearly all small-cell deployments are operator-specific, but in the future, deployments supporting multiple operators could reduce the cost per operator to provide coverage.

A candidate band for neutral-host small cells is 3.5 GHz, using LTE TDD and MulteFire as potential technologies. Wi-Fi technology also addresses neutral-host configurations at the access level, but it has roaming and authentication challenges. HotSpot 2.0 (covered in the appendix) addresses roaming and authentication.

Unlicensed Spectrum Integration

Unlicensed spectrum is becoming ever more important to mobile broadband networks. Initial use was rudimentary offload onto Wi-Fi networks, but now, Wi-Fi networks are becoming more tightly integrated into cellular networks. Efforts are also underway to use LTE in unlicensed spectrum.

Unlicensed spectrum adds to capacity in two ways. First, a large amount of spectrum (approximately 500 MHz) is available across the 2.4 GHz and 5 GHz bands, with the 3.5 GHz band adding further spectrum in the future. A significant amount of unlicensed spectrum is also available in mmWave bands, with 7 GHz already in use in the United States (57 to 64 GHz) and an additional 7 GHz in 5G spectrum allocations. Second, unlicensed spectrum is mostly used in small coverage areas, resulting in high-frequency re-use and much higher throughput rates per square meter of coverage versus typical cellular deployments.

The IEEE 802.11 family of technologies has experienced rapid growth, mainly in private deployments. The latest 802.11 standard, 802.11ac, offers peak theoretical throughputs in excess of 1 Gbps and improved range through use of higher-order MIMO. 802.11ac Wave 2 products include a Multi-user MIMO capability that further increases capacity and throughput. IEEE is developing a subsequent version, 802.11ax, that emphasizes capacity

---

[50] 5G Americas and Small Cell Forum, *Multi-operator and neutral host small cells; Drivers, architectures, planning and regulation,* December 2016. Report available at http://www.5gamericas.org/files/4914/8193/1104/SCF191_Multi-operator_neutral_host_small_cells.pdf.

improvements as well as higher throughputs. In the mmWave frequencies, IEEE has developed 802.11ad, which operates at 60 GHz, and the standards body is currently working on a successor technology, 802.11ay.

Integration between mobile broadband and Wi-Fi networks can be either loose or tight. Loose integration means data traffic routes directly to the internet and minimizes traversal of the operator network. This is called "local breakout." Tight integration means data traffic, or select portions thereof, may traverse the operator core network. An example is Wi-Fi calling, which uses IP Multimedia Subsystem.

Although offloading onto Wi-Fi can reduce traffic on the core network, the Wi-Fi network does not necessarily always have greater spare capacity than the cellular network. The goal of future integrated cellular/Wi-Fi networks is to intelligently load balance between the two. Simultaneous cellular/Wi-Fi connections will also become possible. For example, in Release 13, 3GPP introduces link aggregation of Wi-Fi and LTE through LWA and LWIP.

Successfully offloading data and providing users a good experience mandates measures such as automatically provisioning subscriber devices with the necessary Wi-Fi configuration options and automatically authenticating subscribers on supported public Wi-Fi networks. Many stakeholders are working toward tighter integration between Wi-Fi and cellular networks.

Another approach for using unlicensed spectrum employs LTE as the radio technology, initially in a version referred to as LTE-Unlicensed, which works with Releases 10-12 of LTE, as defined in the LTE-U Forum. In Release 13, 3GPP specified LAA, which implements listen-before-talk capability, a requirement for unlicensed operation in Europe and Japan. Initially, carrier aggregation combines a licensed carrier with an unlicensed 20 MHz carrier in the 5 GHz band as a supplemental channel, with the aggregation of multiple unlicensed channels an eventual possibility. Operating LTE in unlicensed bands could decrease the need for handoffs to Wi-Fi. Up to 32 unlicensed carriers (of 20 MHz each) can be aggregated to theoretically access 640 MHz of unlicensed spectrum. LAA may also be deployed in 3.5 GHz bands. Enhanced LAA (eLAA), specified in Release 14, adds uplink use of unlicensed spectrum. Carriers are now conducting LAA trials.[51]

A concern with using LTE in unlicensed bands is whether it will be a fair neighbor to Wi-Fi users. LTE-U based on Release 10-12 addresses this concern by selecting clear channels to use and measuring the channel activity of Wi-Fi users, then using an appropriate duty cycle for fair sharing. License-Assisted Access in Release 13 adds listen-before-talk (LBT) and implements other regulatory requirements that exist in some countries. 3GPP conducted a study and concluded that, "A majority of sources providing evaluation results showed at least one LBT scheme for LAA that does not impact Wi-Fi more than another Wi-Fi network."[52]

To address co-existence, the cellular industry worked with the Wi-Fi Alliance in 2016 to develop a test plan for LTE-U. The testing goal was to verify that, in a laboratory

---

[51] For example, see AT&T, "AT&T Reaches Wireless Speeds of More than 750 Mbps with LTE Licensed Assisted Access (LTE-LAA) Field Trials," available at http://about.att.com/story/lte_licensed_assisted_access_field_trials.html.

[52] 3GPP, *Technical Specification Group Radio Access Network; Study on Licensed-Assisted Access to Unlicensed Spectrum; (Release 13). 36.889.* See section 9, "Conclusions."

environment, an LTE-U base station does not impact a Wi-Fi network any more than another Wi-Fi access point.[53]

MulteFire, specified by the MulteFire Alliance, is an application of LTE in unlicensed bands that does not require an anchor in licensed spectrum, opening up the possibility of deployments by non-operator entities, including internet service providers, venue operators, and enterprises. Under a roaming arrangement with cellular operators, LTE customers could roam into MulteFire networks. Figure 24 shows the evolution of the different versions of LTE for unlicensed bands.
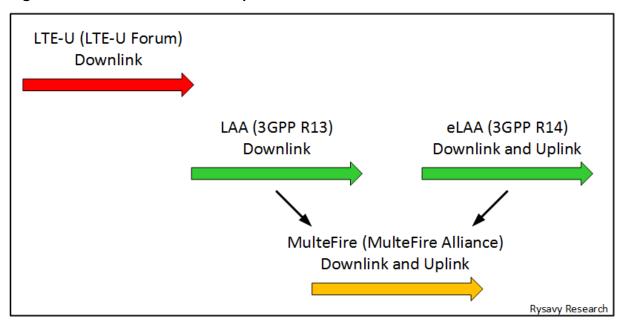
Figure 24: Timeline Relationship of LTE-U, LAA, eLAA, and MulteFire

An alternative approach for integrating Wi-Fi is LWA. LTE handles the control plane, but connections occur over separate LTE base stations and Wi-Fi access points. LWA benefits operators that wish to emphasize Wi-Fi technology for harnessing capacity in unlicensed spectrum. LWIP is a variation of LWA that also integrates LTE and Wi-Fi, but by integrating at a higher level of the protocol stack (IP instead of PDCP), it facilitates use of existing Wi-Fi equipment and devices, with integration typically occurring at the eNodeB.

Figure 25 shows how the different technologies exploit licensed and unlicensed spectrum.

---

[53] See Wi-Fi Alliance, "Unlicensed Spectrum," http://www.wi-fi.org/discover-wi-fi/unlicensed-spectrum.
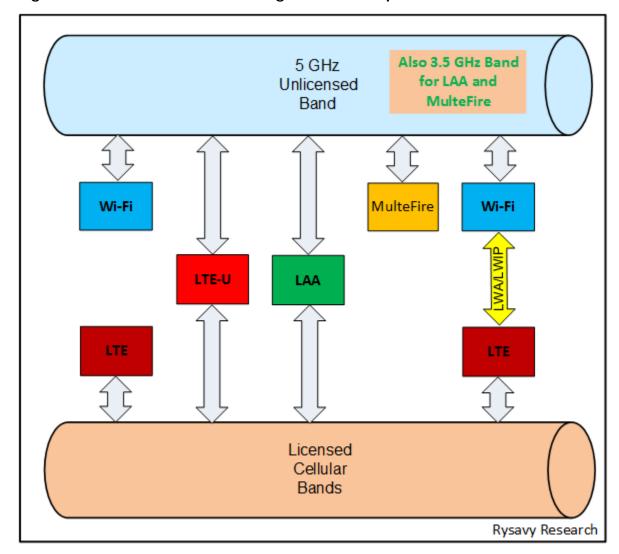
Figure 25: How Different Technologies Harness Spectrum

Table 9 summarizes the different uses of unlicensed spectrum for public mobile broadband networks.

Table 9: Approaches for Using Unlicensed Spectrum.

	Technology	Attributes
Wi-Fi	Ever-more-sophisticated means to integrate Wi-Fi in successive 3GPP Releases.	Combining Wi-Fi with cellular increases capacity.
Release 13 RAN Controlled LTE WLAN Interworking	Base station can instruct the UE to connect to a WLAN for offload.	Available in late 2017 or 2018 timeframe.

Release 10-12 LTE-U Based on LTE-U Forum Specifications	LTE-U Forum-specified approach for operating LTE in unlicensed spectrum.	Available in 2017. More seamless than Wi-Fi. Cannot be used in some regions (e.g., Europe, Japan).
Release 13 Licensed-Assisted Access	3GPP-specified approach for operating LTE in unlicensed spectrum. Downlink only.	Available in late 2017 or 2018 timeframe. Designed to address global regulatory requirements.
Release 14 Enhanced Licensed-Assisted Access	Addition of uplink operation.	Available in 2019 or 2020 timeframe.
5G Unlicensed Operation	Release 15 has a study item for approaches that Release 16 will standardize.	Potentially available in 2021.
MulteFire	Does not require a licensed anchor.	Potentially creates a neutral-host small cell solution.
LWA	Aggregation of LTE and Wi-Fi connections at PDCP layer.	Part of Release 13. Available in late 2017 or 2018 timeframe.
LWIP	Aggregation of LTE and Wi-Fi connections at IP layer.	Part of Release 13. Available in late 2017 or 2018 timeframe.

Cellular operators are currently emphasizing simple offload to Wi-Fi or LTE-U/LAA. Aggregation techniques, such as LWA and LWIP, do not currently have market traction.

Refer to the appendix section "Unlicensed Spectrum Integration" for further technical details.

Internet of Things and Machine-to-Machine

Machine-to-machine communications, now evolving into the Internet of Things, is a vast opportunity for wireless communications, with all 3GPP technologies potentially playing roles.

The lowest-cost cellular devices enabling M2M communications today are GPRS modems, which risk becoming obsolete as operators sunset their GSM systems. HSPA is also used for M2M communications, as is LTE, which has been optimized to efficiently communicate small bursts of information, making it particularly well suited for M2M.

Low-cost GSM (through Enhanced Coverage GSM IoT [EC-GSM-IoT]) and LTE modem options in 3GPP releases 10 through 13 reduce cost, improve communications range, and extend battery life. See the appendix section "Internet of Things and Machine-to-Machine" for details.

3GPP had a Release 13 study item on how LTE technologies can operate for vehicle communications, including vehicle-to-vehicle and vehicle-to-infrastructure, leveraging

device-to-device communications capabilities already specified for LTE in Releases 12 and 13.[54] This work is now being standardized in Release 14.

Developers will use 3GPP wireless technologies for many IoT applications. In other instances, developers will use local area technologies, such as Wi-Fi, Bluetooth Low Energy, and ZigBee. New Low-Power Wide-Area (LPWA) wireless technologies emerging specifically to support IoT include Ingenu, LoRa, and Sigfox. The low-power operation of some of these technologies, including LTE, will permit battery operation over multiple years. Table 10 summarizes the various technologies.

Table 10: Wireless Networks for IoT

Technology	Coverage	Characteristics	Standardization/ Specifications
GSM/GPRS/EC-GSM-IoT	Wide area. Huge global coverage.	Lowest-cost cellular modems, risk of network sunsets. Low-throughput.	3GPP
HSPA	Wide area. Huge global coverage.	Low-cost cellular modems. Higher power, high throughput.	3GPP
LTE, NB-IoT	Wide area. Increasing global coverage.	Wide area, expanding coverage, cost/power reductions in successive 3GPP releases. Low to high throughput options.	3GPP
Wi-Fi	Local area.	High throughput, higher power.	IEEE
ZigBee	Local area.	Low throughput, low power.	IEEE
Bluetooth Low Energy	Personal area.	Low throughput, low power.	Bluetooth Special Interest Group
LoRa	Wide area. Emerging deployments.	Low throughput, low power. Unlicensed bands (sub 1 GHz, such as 900 MHz in the U.S.)	LoRa Alliance[55]
Sigfox	Wide area. Emerging deployments.	Low throughput, low power. Unlicensed bands (sub 1 GHz such as 900 MHz in the U.S.)	Sigfox[56]

---

[54] 3GPP, *3GPP TR 36.885, Technical Specification Group Radio Access Network; Study on LTE-based V2X Services; (Release 14).*

[55] For details, see LoRa Alliance, https://www.lora-alliance.org/.

[56] For details, see Sigfox, https://www.sigfox.com/en.

Technology	Coverage	Characteristics	Standardization/ Specifications
Ingenu (previously OnRamp Wireless)	Wide area. Emerging deployments.	Low throughput, low power. Using 2.4 GHz ISM band. Uses IEEE 802.15.4k.	Ingenu[57]
Weightless	Wide area. Expected deployments.	Low throughput, low power. Unlicensed bands (sub 1 GHz such as TV White-Space and 900 MHz in the U.S.)	Weightless Special Interest Group[58]

Security is of particular concern to both developers and users of IoT technology. An increasing amount of network-connected infrastructure will result in new security vulnerabilities that are being addressed by concerted effort from the industry.[59]

Smart Antennas and MIMO

Smart antennas, defined with progressively greater capabilities in successive 3GPP releases, provide significant gains in throughput and capacity. By employing multiple antennas at the base station and the subscriber unit, the technology either exploits signals traveling through multiple paths in the environment or does beam steering, in which multiple antennas coordinate their transmissions to focus radio energy in a particular direction.

Initial low-band LTE deployments used 2X2 MIMO on the downlink (two base station transmit antennas, two mobiles receive antennas) and 1X2 on the uplink (one mobile transmit antenna, two base station receive antennas). In the higher bands, 2X2 downlink MIMO has been deployed, but it is more common to employ four antennas for uplink reception in a 1X4 configuration. LTE deployments are now using 4X2 MIMO and 4X4 MIMO on the downlink (four base station transmit antennas). LTE specifications encompass higher-order configurations, such as 4X4 MIMO, 8X2 MIMO, and MU-MIMO on the downlink and 1X4 on the uplink. Practical considerations, such as antenna sizes that are proportional to wavelength, dictate MIMO options for different bands.

Engineers are now experimenting with what are called massive MIMO systems, which employ a far larger number of antenna elements at the base station—64, 128, and eventually even more. Use in 5G of cmWave and mmWave bands, with their short wavelengths, will facilitate massive MIMO, but even before then, 3GPP is developing specifications for massive MIMO for 4G systems in what it calls full-dimension MIMO (FD-MIMO). Release 14 specifies configurations with up to 32 antennas at the base station.

---

[57] For details, see Ingenu, https://www.ingenu.com/.

[58] For details, see http://www.weightless.org/.

[59] For further insight, refer to the Ericsson white paper, *IoT Security,* February 2017, available at https://www.ericsson.com/assets/local/publications/white-papers/wp-iot-security-february-2017.pdf.

Massive MIMO is practical even in cellular frequencies. For example, at 2.5 GHz, an 8X8 array using half wavelength spacing would produce a form factor of 50 cm X 50 cm. Applications of such arrays include beamforming along a horizontal direction as well as beamforming in a vertical direction, such as to serve different levels of high-rise buildings.

See the appendix section "LTE Smart Antennas" and "LTE-Advanced Antenna Technologies" for further details.

Virtualization

Virtualization refers to implementing the functions of infrastructure nodes in software on commercial "off-the-shelf" computing equipment. The approach promises lower capital expenditures, lower operating costs, faster deployment of new services, energy savings, and improved network efficiency. With NFV, multiple tenants will be able to share the same infrastructure, facilitating, for example, mobile virtual network operator (MVNO) and multi-operator virtualized RAN arrangements. NFV, however, also constitutes an entirely new way of building and managing networks, so widespread adoption will occur over a long period.

Both the core network and portions of the radio-access network can be virtualized. The core network, consisting of fewer nodes, is an easier starting point. Virtualizing RAN elements, although more complex, could eventually provide the greatest network efficiency gains, particularly for small-cell deployments where it can facilitate coordination among cells and use of methods such as CoMP and interference coordination. Unlike the core, virtualizing the entire RAN is not possible because a Physical Network Function must terminate the radio interface. As operators virtualize their core networks, they put in place the systems and know-how to extend virtualization to the RAN.

The European Telecommunications Standards Institute (ETSI) is standardizing a framework, including interfaces and reference architectures for virtualization. Other standards and industry groups involved include 3GPP, the Open Networking Foundation, OpenStack, OpenDaylight, and OPNFV.

Figure 26 shows the ETSI framework, in which virtualized network functions are the nodes or applications by which operators build services.

Figure 26: ETSI NFV High-Level Framework

Some specific use cases for NFV include:

❑ **5G.** 5G networks will likely be fully virtualized.

❑ **IMS and VoLTE.** IMS is necessary for VoLTE, but an NFV approach could reduce the complexity associated with the multiple nodes and interfaces in the IMS architecture.

❑ **Virtualized EPC (VEPC).** The Evolved Packet Core, consisting of the Serving Gateway (SGW), the Packet Gateway (PGW), and Mobile Management Entity (MME), can be virtualized, but doing so will require meeting operator bandwidth, latency, and control plane service requirements.

❑ **New VEPC Services.** With a virtualized EPC, an operator can more easily create MVNO services, each with its own virtualized MME, SGW, and PGW. An M2M virtualized service is another example of offering a more finely tuned service for the target application. Because the PGW connects to external networks, further opportunities exist for virtualized services to augment networking functions, including video caching, video optimization, parental controls, ad insertion, and firewalls.

❑ **Cloud RAN.** Pooling of baseband processing in a cloud RAN can, but does not necessarily, use virtualization techniques. Separating the radio function from baseband processing typically requires transporting digitized radio signals across high-bandwidth (multi-Gbps) fiber connections, sometimes referred to as

fronthauling. Refer to the appendix section "Cloud Radio-Access Network (RAN) and Network Virtualization" for a more detailed technical discussion.[60]

Because of higher investment demands, RAN virtualization will take longer to deploy than core network virtualization and likely will occur selectively for small-cell deployments.

For additional details, refer to the 5G Americas white paper, *Bringing Network Function Virtualization to LTE*.[61]

Multi-Access Edge Computing

ETSI is standardizing Multi-access Edge Computing, previously known as Mobile-Edge Computing, a technology that empowers a programmable application environment at the edge of the network, within the RAN.[62] Goals include reduced latency, more efficient network operation for certain applications, and an improved user experience. Although MEC emphasizes 5G, it can also be applied to 4G LTE networks.

Applications that will benefit are ones that require server-side processing but are location specific. Examples include:

❑ Augmented reality.

❑ Intelligent video processing, such as transcoding, caching, and acceleration.

❑ Connected cars.

❑ Premise-based IoT gateways.

Multicast and Broadcast

Another important new service is video streaming via multicast or broadcast functions. 3GPP has defined multicast/broadcast capabilities for both HSPA and LTE. Mobile TV services have experienced little business success so far, but broadcasting uses the radio resource much more efficiently than having separate point-to-point streams for each user. For example, users at a sporting event might enjoy watching replays on their smartphones. The technology supports these applications; it is a matter of operators and content providers finding appealing applications. The appendix covers technical aspects in more detail.

Remote SIM Provisioning

The GSM Association (GSMA) is developing specifications that make it possible for consumers to purchase unprovisioned devices, select the operator of their choice and then download the subscriber identity module (SIM) application into the device.[63] This capability

---

[60] For further details, see "Network Functions Virtualisation," http://www.etsi.org/technologies-clusters/technologies/nfv. Viewed May 17, 2017.

[61] Available at http://www.4gamericas.org/files/1014/1653/1309/4G_Americas_-_NFV_to_LTE_-_November_2014_-_FINAL.pdf.

[62] For further details, see ETSI, "Multi-access Edge Computing," http://www.etsi.org/technologies-clusters/technologies/multi-access-edge-computing. Accessed May 17, 2017.

[63] For details, see GSMA, "A New SIM," available at https://www.gsma.com/rsp/, viewed June 8, 2017.

benefits devices such as watches, health bands, health monitors, and other small connected items. Initial products may be available in 2017.

VoLTE, RCS, WebRTC, and Wi-Fi Calling

Voice has evolved from a separate circuit-switched service in 2G and 3G networks to a packet-switched service in 4G LTE networks that can integrate with other services and applications, such as messaging and video calling. Elements that make these capabilities possible include the quality-of-service mechanisms in LTE, the IMS platform discussed above, implementation of Rich Communications Suite, compliance with GSMA IR.92 guidelines, and optional support for WebRTC.

Voice Support and VoLTE

While 2G and 3G technologies were deployed from the beginning with both voice and data capabilities, LTE networks can be deployed with or without voice support. Moreover, there are two methods available: circuit-switched fallback (CSFB) to 2G/3G and VoIP. Most operators deploying LTE used CSFB initially but are now migrating to VoIP methods with VoLTE that uses IMS. Initial VoLTE deployments occurred in 2012. Because VoLTE needs new software in phones, the transition from circuit-switched voice to VoLTE on a large scale will occur over a number of years as users upgrade their devices.

For the time being, 3GPP operators with UMTS/HSPA networks will continue to use circuit-switched voice for their 3G connections.

Using VoLTE, operators can offer high-definition (HD) voice using the new Adaptive Multi-Rate Wideband (AMR-WB) voice codec. HD voice not only improves voice clarity and intelligibility, it suppresses background noise. AMR-WB extends audio bandwidth to 50-7000 Hz compared with the narrowband codec that provides audio bandwidth of 80-3700 Hz. HD voice will initially function only between callers on the same network. 3GPP has also developed a new voice codec, called "Enhanced Voice Services" (EVS), which will be the successor to AMR and AMR-WB codecs.

Other advantages of LTE's packetized voice include being able to combine it with other services, such as video calling and presence; half the call set-up time of a 3G connection; and high voice spectral efficiency. With VoLTE's HD voice quality, lower delay, and higher capacity, operators can compete against OTT VoIP providers. Due to traffic prioritization, VoLTE voice quality remains high even under heavy loads that cause OTT-voice service to deteriorate.

Applications based on WebRTC will also increasingly carry voice sessions. See the section "VoLTE and RCS" in the appendix for more details on LTE voice support.

Rich Communications Suite

An initiative called "Rich Communications Suite" (RCS), supported by many operators and vendors, builds on IMS technology to provide a consistent feature set as well as implementation guidelines, use cases, and reference implementations. RCS uses existing standards and specifications from 3GPP, Open Mobile Alliance (OMA), and GSMA and enables interoperability of supported features across operators that support the suite. RCS supports both circuit-switched and packet-switched voice and can interoperate with LTE packet voice.

Core features include:

❑ A user capability exchange or service discovery with which users can know the capabilities of other users.

❑ Enhanced (IP-based) messaging (supporting text, IM, and multimedia) with chat and messaging history.

❑ Enriched calls that include multimedia content (such as photo or video sharing) during voice calls. This could become the primary way operators offer video calling.

The primary drivers for RCS adoption are the ability to deploy VoLTE in a well-defined manner and to support messaging in the IP domain. RCS addresses the market trend of users moving away from traditional text-based messaging and provides a platform for operator-based services that compete with OTT messaging applications. Figure 27 shows the evolution of RCS capability, including the addition of such features as messaging across multiple devices, video calling, video sharing, and synchronized contact information across multiple devices.

Figure 27: Evolution of RCS Capability.[64]

WebRTC

WebRTC is an open project supported by Google, Mozilla, and Opera within the Internet Engineering Taskforce (IETF) that enables real-time communications in Web browsers via JavaScript APIs. 3GPP Release 12 specifications define how WebRTC clients can access IMS services, including packet voice and video communication. WebRTC operating over IMS gains the additional benefit of seamless transition across transport networks, for example, LTE to Wi-Fi.

Operators can integrate WebRTC with RTC, facilitating development of vertical applications such as telemedicine and customer service. WebRTC and RCS are more complementary

---

[64] 4G Americas, *VoLTE and RCS Technology - Evolution and Ecosystem*, Nov. 2014.

than competitive. Both, through application interfaces, can provide access to underlying network functions.

Wi-Fi Calling

Another advantage of the VoLTE/IMS/RCS architecture is that it is agnostic to the user connection, meaning voice and video service can extend to Wi-Fi connections as easily as LTE connections. Wi-Fi calling can be advantageous in coverage areas were the Wi-Fi signal has better quality than an LTE signal. For video calling, use of Wi-Fi will also reduce data consumption over the cellular connection. By implementing a standards-based approach, as opposed to OTT-voice approaches, called parties see the same phone number regardless of network and can reach the subscriber using that phone number.

Previous technical approaches, such as Generic Access Network (GAN, initially called Unlicensed Mobile Access [UMA]), did not include as robust a handover mechanism as is provided by VoLTE/IMS.

For the best-quality voice in a Wi-Fi network, the device and Wi-Fi network should implement Wi-Fi Multimedia (WMM), which gives voice packets higher priority than other data traffic. WMM is especially necessary in congested networks. In addition, the Access Network Discovery and Selection Function (ANDSF) and cellular-WLAN enhancement features in 3GPP Release 12 have policies for enabling voice handover between LTE and Wi-Fi.

Roaming with Wi-Fi calling will need to address whether the visited network's IMS infrastructure handles the Wi-Fi call or whether the home network's IMS does.

Public Safety

An important LTE application is for public safety, initially as a broadband data service and eventually for mission-critical voice service. Current public safety networks use technologies, such as Terrestrial Trunked Radio (TETRA) in Europe and Project 25 (P25) in the United States, that provide mission-critical voice but only narrowband data.

In the United States, the government has made 20 MHz of spectrum available at 700 MHz in Band 14 and created the First Responder Network Authority (FirstNet), an independent authority within the National Telecommunications and Information Administration (NTIA) to provide a nationwide public-safety broadband network. AT&T will build and deploy the network.[65]

Another country driving the use of LTE for public safety is the United Kingdom, where the UK Home Office has a program for the Emergency Service Network.

Using LTE for public safety is a complex undertaking because public-safety needs differ from those of consumers. Addressing these needs requires both different features, which 3GPP is incorporating in multiple releases of LTE specifications, and different network deployment approaches. Public safety also has different device-level needs than consumers.

LTE Features for Public Safety

Some broadband applications for public safety can use standard LTE capability. For example, sending email, accessing a database, or streaming a video may not require any special features. Other applications, however, require new capabilities from 3GPP standards, including:

Group Communication

Available in Release 12, the Group Communication Service (GCS) application server, using one-to-one (unicast) and one-to-many communications (broadcast), will be able to send voice, video, or data traffic to multiple public-safety devices. The broadcast mode will employ eMBMS to use radio resources efficiently, but if coverage is weak, a unicast approach may deliver data more reliably. The system will be able to dynamically switch between broadcast and unicast modes. Release 14 adds single-cell point-to-multipoint transmission.

Proximity-Based Services (Device-to-Device)

With proximity-based services, defined in Release 12, user devices can communicate directly, a capability that benefits both consumers and public safety. This type of communication is called sidelink communication. Consumer devices can find other devices only with assistance from the network, but for public safety, devices will be able to communicate directly with other devices independently of the network.

With Release 13, devices will be able to act as relays for out-of-coverage devices, such as inside a building.

---

[65] For details, see "FirstNet AT&T," https://www.corp.att.com/public-safety/att-firstnet/. See also RCR Wireless, *Editorial Report: Public Safety LTE*, March 2017. Available at http://content.rcrwireless.com/20170322_Public_Safety_Report.

The appendix section "Proximity Services (Device-to-Device)" discusses this feature in greater detail.

Mission-Critical Push-to-Talk

MCPTT, being defined in Release 13, provides one-to-one and one-to-many push-to-talk communications services. With this feature, available in the 2018 timeframe, public-safety organizations will be able to consider retiring legacy voice-based systems.

Mission-Critical Video over LTE and Mission-Critical Data over LTE

Release 14 adds Mission-Critical Video over LTE and Mission-Critical Data over LTE, designed to work with Mission-Critical Push-to-Talk, giving first responders more communications options.

Prioritization

To prevent interference with public-safety operations in emergency situations experiencing high load, the network can prioritize at multiple levels. First, the network can bar consumer devices from attempting to access the network, thus reducing signaling load. Second, the network can prioritize radio resources, giving public-safety users higher priority. Third, using a new capability called "Multimedia Priority Service" (MPS), the network can prioritize a connection between an emergency worker and a regular subscriber. Finally, the network can assign specific quality-of-service (QoS) parameters to specific traffic flows, including guaranteed bit rate. 3GPP has defined specific QoS quality-class identifiers for public safety.

High Power

Release 11 defines higher power devices for the public safety band that can operate at 1.25 Watts, improving coverage and reducing network deployment costs.

Isolated operation

With Release 13, a base station can continue offering service even with the loss of backhaul, a capability that will benefit public-safety personnel in disaster situations.

Relays

Figure 28 summarizes the more than eighteen features in 3GPP relays that apply to public safety.

Figure 28: Summary of 3GPP LTE Features to Support Public Safety[66]

3GPP Rel-8	3GPP Rel-9	3GPP Rel-10	3GPP Rel-11	3GPP Rel-12	3GPP Rel-13
• Mobile data connections • Basic support for Voice over LTE (telephony) • Support for LTE Band 14 • a rich set of QoS priority and pre-emption features • Highly secure authentication and ciphering	• Location services and positioning support for LTE • Multimedia Broadcast / Multicast Service • E911 or emergency calling support • Enhanced Home LTE base station: "Cell On Wheels" • Self-Organizing Networks (SONs)	• Physical layer enhancements to increase data throughput (including LTE-Advanced features) • Relays for LTE, e.g. to allow a base station mounted on a fire vehicle to relay communications from firefighters in a basement back to the network.	• High power devices for Band 14 - 1.25 Watts for public safety devices significantly improving the coverage of an LTE network, benefiting public safety users and reducing network deployment costs.	• Group Communication System Enablers for LTE • Proximity-based Services 3GPP work ongoing - completion expected 1Q2015	• Mission Critical Push-to-Talk • Enhancements to Proximity-based Services • Isolated E-UTRAN Operation for Public Safety • MBMS Enhancements 3GPP work started - completion expected 2016

Deployment Approaches

Because huge infrastructure investments would be required for a network dedicated solely to public safety, industry and government are evaluating approaches in which public-safety uses can leverage existing commercial network deployments. One caveat is that public-safety networks have more stringent resilience and security requirements than commercial networks.

Shared Network

As depicted in Figure 29, multiple sharing approaches are possible:

1. In this scenario, a public-safety entity owns and operates the entire network, an approach that gives public-safety organizations the greatest control over the network but at the highest cost.

2. A commercial operator shares its radio-access network, including cell sites and backhaul, but the public-safety entity manages core network functions including gateways, the Mobile Management Entity, the Home Subscriber Server (HSS), and public-safety application servers. Because the radio-access network is the costliest part of the network, this approach significantly reduces the amount public safety must invest in the network. Even though the RAN is shared, public safety still can use its dedicated spectrum.

3. In an MVNO approach, the operator shares its cell sites and backhaul as well as some core network functions, such as the MME and Serving Gateway. Public safety manages a small number of network functions, such as the Packet Gateway, HSS, and its application servers.

---

[66] Nokia, *LTE networks for public safety services*, 2014. Available at http://networks.nokia.com/sites/default/files/document/nokia_lte_for_public_safety_white_paper.pdf.

---

4. A final approach, not shown in the figure, is one in which the mobile operator hosts all of the elements shown in the figure and public safety manages only its application servers.

Figure 29: Sharing Approaches for Public-Safety Networks

Resilience

Public safety may need greater resilience than found in commercial networks, including hardware redundancy, geographic redundancy, load balancing, fast re-routing in IP networks, interface protection, outage detection, self-healing, and automatic reconfiguration.

Security

Public-safety networks may have higher security requirements than commercial networks, including physical security of data centers, core sites, and cell sites. Whereas LTE networks do not have to encrypt traffic in backhaul and core networks, public-safety applications may choose to encrypt all IP traffic using virtual private networking approaches.

Coverage

A number of approaches can ensure the broadest possible coverage for public-safety networks. First, public-safety frequencies at 700 MHz already propagate and penetrate well. Next, public-safety devices will be able to transmit at higher power. In addition, base stations can employ four-way receiver diversity and higher-order sectorization. For disaster situations, public safety can also use rapidly deployable small cells, such as on trailers. Finally, proximity-based services operating in a relay mode, as discussed above, can extend coverage.

Device Considerations for Public Safety

Public-safety devices will have unique requirements, including guaranteed network access under all conditions and guidelines for how devices are shared among users.

Access to Commercial Networks

Public-safety devices could be designed to also communicate on commercial LTE networks, providing an alternative communications avenue when the device cannot connect to a public-safety network. Subscriptions to all major commercial networks would make this approach the most effective. Wi-Fi capability further extends this concept.

Device Sharing

Because public-safety devices may be shared among personnel, user profiles cannot be stored on USIM cards stored in the devices. Bluetooth-based remote SIMs are one approach to address this problem.

Expanding Capacity

Wireless technology is playing a profound role in networking and communications, even though wireline technologies such as fiber have inherent capacity advantages.

Over time, wireless networks will gain substantial additional capacity through the methods discussed in the next section. While they will compete with copper twisted pair and coax, they will never catch up to fiber. The infrared frequencies used in fiber-optic communications have far greater bandwidth than radio. As a result, one fiber-optic strand has greater bandwidth than the entire usable radio spectrum to 100 GHz, as illustrated in Figure 30.[67]
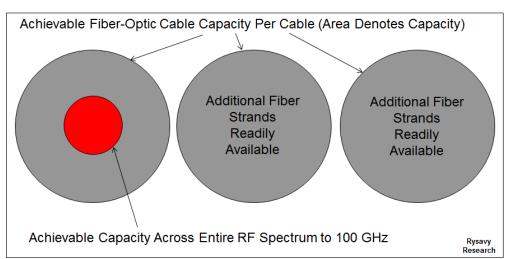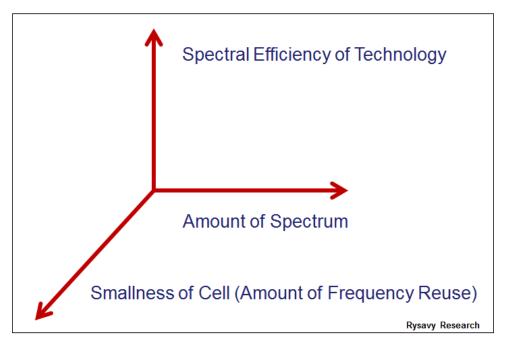
Figure 30: RF Capacity vs. Fiber-Optic Cable Capacity

A dilemma of 4G mobile broadband is that it *can* provide a broadband experience similar to wireline, but it *cannot* do so for all subscribers in a coverage area at the same time. Hence, operators must carefully manage capacity, demand, policies, pricing plans, and user expectations. Similarly, application developers must become more conscious of the inherent constraints of wireless networks. 5G, with its far greater capacity, at least in small deployments, will increase the percentage of users that use wireless connections as their only broadband.

Three factors determine wireless network capacity, as shown in Figure 31: the amount of spectrum, the spectral efficiency of the technology, and the size of the cell. Because smaller cells serve fewer people in each cell and because there are more of them, small cells are a major contributor to increased capacity.

---

[67] One fiber-optic cable can transmit over 10,000 Gbps compared with all wireless spectrum to 100 GHz, which, even at an extremely high spectral efficiency 10 bps/Hz, would have only 1,000 Gbps of capacity.

Figure 31: Dimensions of Capacity

Given the relentless growth in usage, mobile operators are combining multiple approaches to increase capacity and managing congestion:

❑ **More spectrum.** Spectrum correlates almost directly to capacity, and more spectrum is becoming available globally for mobile broadband. In the U.S. market, the 2010 FCC National Broadband Plan is in the process of adding an additional 500 MHz of spectrum by 2020. mmWave band spectrum for 5G will provide far more spectrum, but propagation characteristics will restrict its use to small cells. Multiple papers by Rysavy Research and others[68] argue the critical need for additional spectrum.

❑ **Unpaired spectrum.** LTE TDD operates in unpaired spectrum. In addition, technologies such as HSPA+ and LTE permit the use of different amounts of spectrum between downlink and uplink. Additional unpaired downlink spectrum can be combined with paired spectrum to increase capacity and user throughputs.

❑ **Supplemental downlink.** With downlink traffic five to ten times greater than uplink traffic, operators often need to expand downlink capacity rather than uplink capacity. Using carrier aggregation, operators can augment downlink capacity by combining separate radio channels.

❑ **Spectrum sharing.** Policy makers are evaluating how spectrum might be shared between government and commercial entities. Although a potentially promising approach for the long term, sharing raises complex issues, as discussed further in the section "Spectrum Developments."

❑ **Increased spectral efficiency.** Newer technologies are spectrally more efficient, meaning greater aggregate throughput using the same amount of spectrum. Wireless technologies such as LTE, however, are reaching the theoretical limits of spectral

---

[68] See multiple papers on spectrum and capacity at http://www.rysavy.com/writing.

efficiency, and future gains will be quite modest, allowing for a possible doubling of LTE efficiency over currently deployed versions. See the section "Spectral Efficiency" for a further discussion.

❑ **Smart antennas.** Through higher-order MIMO and beamforming, smart antennas gain added sophistication in each 3GPP release and are the primary contributor to increased spectral efficiency (bps/Hz). Massive MIMO, beginning in Release 13, will support 16-antenna-element systems and in 5G, will expand to potentially hundreds of antenna elements.

❑ **Uplink gains combined with downlink carrier aggregation.** Operators can increase network capacity by applying new receive technologies at the base station (for example, large-scale antenna systems such as massive MIMO) that do not necessarily require standards support. Combined with carrier aggregation on the downlink, these receive technologies produce a high-capacity balanced network, suggesting that regulators should in some cases consider licensing just downlink spectrum.

❑ **Small cells and heterogeneous networks.** Selective addition of picocells to macrocells to address localized demand can significantly boost overall capacity, with a linear increase in capacity relative to the number of small cells. HetNets, which also can include femtocells, hold the promise of increasing capacity gains by a factor of four and even higher with the introduction of interference cancellation in devices. Distributed antenna systems (DAS), used principally for improved indoor coverage, can also function like small cells and increase capacity. Actual gain will depend on a number of factors, including number and placement of small cells,[69] user distribution, and any small-cell selection bias that might be applied.

❑ **Offload to unlicensed spectrum.** Using unlicensed spectrum with Wi-Fi or LTE operation in unlicensed spectrum offers another means of offloading heavy traffic. Unlicensed spectrum favors smaller coverage areas because interference can be better managed, so spectral re-use is high, resulting in significant capacity gains.

❑ **Higher level sectorization.** For some base stations, despite the more complex configuration involved, six sectors can prove advantageous versus the more traditional three sectors, deployed either in a 6X1 horizontal plane or 3X2 vertical plane.[70]

Strategies to manage demand include:

❑ **Quality of service (QoS) management.** Through prioritization, certain traffic, such as non-time-critical downloads, could occur with lower priority, thus not affecting other active users. Current network neutrality rules, however, may constrain use of traffic prioritization.[71]

❑ **Off-peak hours.** Operators could offer user incentives or perhaps fewer restrictions on large data transfers during off-peak hours.

---

[69] With small-cell range expansion using a large selection bias, small cells can be distributed uniformly.

[70] An example of vertical layering would be a 3X1 layer at ground level and a separate 3X1 layer for higher levels of surrounding buildings.

[71] For a discussion of this issue, see Rysavy Research, *LTE Congestion Management – Enabling Innovation and Improving the Consumer Experience*, January 2015. Available at http://www.rysavy.com/Articles/2015-01-Rysavy-LTE-Congestion-Management.pdf.

Given a goal of increasing capacity by a factor of 1,000, 50X could roughly be achieved through network densification; 10X through more spectrum, including higher frequencies such as mmWave; and 2X by increases in spectral efficiency.

Based on historical increases in the availability of new spectrum, technologies delivering better spectral efficiency, and increases in the number of cell sites, Rysavy Research has calculated that, over the last thirty-year period, aggregate network capacity has doubled every three years. Rysavy Research expects this trend to continue into the future.

Rysavy Research Analysis:

Aggregate Wireless Network Capacity Doubles Every Three Years.

Spectrum Developments

Licensed spectrum scarcity continues to challenge the industry. Tactics to make the best use of this limited resource include deploying technologies that have higher spectral efficiency; adapting specifications to enable operation of UMTS-HSPA and LTE in all available bands; designing both FDD and TDD versions of technology to take advantage of both paired and unpaired bands; designing carrier aggregation techniques in HSPA+ and LTE-Advanced Pro; and deploying as many new cells, large and small, as is economically and technically feasible. Although all of these industry initiatives greatly expand capacity, they do not obviate the need for additional spectrum. 5G technology will be able to employ frequencies not previously used in cellular systems, spanning 6 GHz to 100 GHz.

An important aspect of UMTS-HSPA and LTE deployment is for infrastructure and mobile devices to accommodate the expanding number of available radio bands. The fundamental system design and networking protocols remain the same for each band; only the frequency-dependent portions of the radios must change. As other frequency bands become available for deployment, standards bodies adapt UMTS-HSPA and LTE for these bands as well.

3GPP has specified LTE for operation in many different bands, and initial use will be more fragmented than the four bands (850 MHz, 900 MHz, 1.8 GHz, 1.9 GHz) that enable global roaming on 2G and the additional two bands (1.7 GHz and 2.1 GHz) that enable 3G roaming. In the Americas, LTE roaming may occur in the 1.7/2.1 GHz (AWS) bands, and globally, LTE roaming may occur in the 1.8 and 2.6 GHz bands. Longer term, operators will re-farm spectrum used for 2G and 3G and apply it to LTE. Unfortunately, the process of identifying new spectrum and making it available for the industry is a lengthy one, as shown in Figure 32.
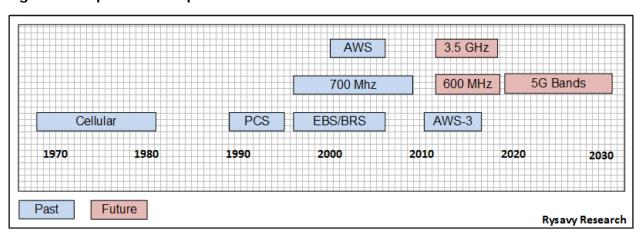
Figure 32: Spectrum Acquisition Time[72]

New short-term spectrum opportunities in the United States include the "small-cell" band from 3550 to 3700 MHz and 5G spectrum.

---

[72] Source for historical data, FCC, *National Broadband Plan*, Chapter 5. Available at http://www.broadband.gov/plan/5-spectrum/, accessed May 18, 2017. Future based on Rysavy Research analysis.

Table 11 summarizes current and future spectrum allocations in the United States.[73]

Table 11: United States Current and Future Licensed Spectrum Allocations

Frequency Band	Amount of Spectrum	Comments
600 MHz	70 MHz	Ultra-High-Frequency (UHF).
700 MHz	70 MHz	Ultra-High Frequency (UHF).
850 MHz	64 MHz	Cellular and Specialized Mobile Radio.
1.7/2.1 GHz	90 MHz	Advanced Wireless Services (AWS)-1.
1695-1710 MHz, 1755 to 1780 MHz, 2155 to 2180 MHz	65 MHz	AWS-3. Uses spectrum sharing.
1.9 GHz	140 MHz	Personal Communications Service (PCS).
2000 to 2020, 2180 to 2200 MHz	40 MHz	AWS-4 (Previously Mobile Satellite Service).[74]
2.3 GHz	20 MHz	Wireless Communications Service (WCS).
2.5 GHz	194 MHz	Broadband Radio Service. Closer to 160 MHz deployable.
	FUTURE	
3.55 to 3.70 GHz	150 MHz	Small-cell band with spectrum sharing and unlicensed use.
Above 6 GHz	Multi GHz	Anticipated for 5G systems beginning in 2017. Based on wavelengths, 3 GHz to 30 GHz is referred to as the cmWave band and 30 GHz to 300 GHz is referred to as the mmWave band.

Today's licensed spectrum networks operate most efficiently and are deployed most cost-effectively using a combination of low-band spectrum, below 1 GHz, for coverage and 1 GHz

---

[73] For international allocations, refer to Wik-Consult, Study for the European Commission, *Inventory and review of spectrum use: Assessment of the EU potential for improving spectrum efficiency*, September 2012. Available at http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/cion_spectrum_inventory_executive_summary_en.pdf.

[74] Supported in 3GPP Band 70, which adds 1995-2000 MHz, pairing it with 1695-1710 MHz in AWS-3 band.

to 3 GHz for capacity. As technology improves, bands in 3 GHz to 100 GHz, and eventually higher, will supplement capacity.

The subsections below provide additional information about the recently completed AWS-3 auction, incentive auctions, the 3.5 GHz, 5G, spectrum harmonization, unlicensed spectrum, and spectrum sharing.

AWS-3

In early 2015, the FCC received close to $45 billion for the U.S. Treasury in the AWS-3 auction, more than twice the amount of any previous auction, demonstrating the value of mid-band spectrum.[75] The auction adds 65 MHz of desirable spectrum to the mobile-broadband industry. The plan is to employ spectrum sharing among commercial networks and select government systems. Eventually, most of these government systems will migrate to another spectrum. 3GPP has specified use of both AWS-3 and AWS-4 spectrum in what it refers to as "Band 66."

This band is asymmetrical, with a downlink of 90 MHz and an uplink of 70 MHz. An operator can use the upper-most 20 MHz of the downlink only with carrier aggregation.
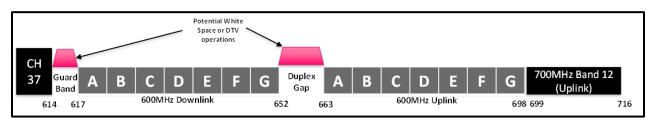
Broadcast Incentive Auction (600 MHz)

The broadcast incentive auction completed in 2017 will reallocate 84 MHz of UHF channels in the 600 MHz band that are currently used by TV broadcasters, with 70 MHz of licensed spectrum and 14 MHz of unlicensed spectrum. The auction was more complicated than past spectrum auctions, when the FCC simply reassigned or designated spectrum for commercial mobile use and then conducted an auction.

In the first stage, the FCC conducted a reverse auction to determine how much spectrum broadcasters might wish to relinquish in exchange for how much compensation. In the second stage, mobile operators bid for spectrum in a forward auction, similar to past spectrum auctions.

Figure 33 shows the final band plan.

Figure 33: 600 MHz Band Plan[76]

Part of the auction process reorganizes and repacks relinquished channels, as well as channels needed for broadcasters that want to keep broadcasting, to make useful blocks of spectrum for mobile broadband. The FCC's goal is to design an auction that will result in

---

[75] For further details, see Rysavy Research, "Latest FCC auction shatters spectrum myths," January 2015. Available at https://gigaom.com/2015/01/17/latest-fcc-auction-shatters-spectrum-myths/.

[76] 5G Americas member contribution.

a uniform nationwide band plan, but in some markets, there may be deviations from that plan.

With a 39-month schedule for winning bidders to move into their new spectrum, the 600 MHz band will be fully available by 2020. However, some operators, will begin using this spectrum in advance of this date, possibly as soon as the second half of 2017, as the spectrum becomes available for use.[77] Operators will use the band for LTE and 5G.

3550 to 3700 MHz "Small-Cell" Band

In the United States, the FCC is in the process of opening the 3550 to 3700 MHz CBRS band. The best use of this band will be small cells and backhaul. The FCC is implementing a three-tier model with incumbent access, priority access with priority access licenses (PALs), and General Authorized Access (GAA) for unlicensed users.[78] Incumbent access will include government radar systems.

Two industry organizations, the Wireless Innovation Forum[79] and the CBRS Alliance,[80] are working for the realization of 3.5 GHz systems.

The mobile broadband industry will mostly use LTE TDD in small-cell configurations, in either a licensed mode or an unlicensed mode using LAA or MulteFire.

Although this band represents a significant amount of new spectrum, potential adopters of the band have expressed concern about proposed rules for use of the band. For example:

❑ IEEE has stated it will not pursue a version of 802.11 for this band because exclusion zones are so large that the percentage of the U.S. population that could be served is below what is required for a successful market.[81]

❑ Licensed users will only be able to obtain three-year licenses with no automatic renewal.[82]

Assuming LAA auctions in early 2018, initial 3.5 GHz deployments could occur in 2018. Because an Environmental Sensing Capability (ESC) has not yet been developed and

---

[77] For example, see FierceWireless, "T-Mobile confirms speedy 600 MHz rollout for 2017, covering 1 million+ square miles," August 8, 2017, available at http://www.fiercewireless.com/node/334911.

[78] For further details, see Official FCC Blog, "Innovation in the 3.5 GHz Band: Creating a New Citizens Broadband Radio Service," March 2015, available at http://www.fcc.gov/blog/innovation-35-ghz-band-creating-new-citizens-broadband-radio-service. See also FCC, *Further Notice of Proposed Rulemaking-- Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550- 3650 MHz Band*, April 23, 2014.

[79] See http://www.wirelessinnovation.org/.

[80] See https://www.cbrsalliance.org/.

[81] IEEE 802.11, *Comments of IEEE 802.11, In the Matter of Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550-3650 MHz Band,"* Jul, 2015. Available at http://apps.fcc.gov/ecfs/document/view?id=60001115064.

[82] See Rysavy article about this band, "Scary Experimentation at 3.5 GHz," Jun. 2016, available at http://www.rysavy.com/Articles/2016-06-Scary-Experimentation-3-5-GHz.pdf. See also AT&T ex-parte FCC communication, http://apps.fcc.gov/ecfs/document/view?id=60001569272.

approved, these initial deployments may not come to fruition in coastal areas where the ESC is needed for co-existence with military systems using those frequencies.

See the section "Spectrum Sharing" for further details of this band.

2.5 GHz Band

In the United States, the primary operator at 2.5 GHz is Sprint. In 2017, Sprint announced that they will deploy 5G in its 2.5 GHz spectrum as soon as the second half of 2019.[83] Having an average of 160 MHz of spectrum at 2.5 GHz in its top 100 markets, Sprint is in the process of deploying LTE-Advanced. In addition, Sprint is in an LTE deployment mode that includes the following features:

❑ Carrier aggregation capability between 2.5 GHz and lower bands (800MHz and 1900 MHz) and aggregating three carriers in a single band.

❑ TDD operation at 2.5 GHz and FDD operation in lower bands.

❑ 8 transmit and 8 receive radios at the base station that enables 4X2 MIMO in combination with beamforming.[84]

5G Bands

As radio technology progresses, it can handle higher frequencies, and it occupies greater bandwidth. 1G systems used 30 kHz radio carriers, 2G in GSM uses 200 kHz carriers, 3G in UMTS uses 5 MHz carriers, and 4G in LTE uses carriers of up to 640 MHz through carrier aggregation.

3GPP is specifying 5G NR to be band agnostic. 5G will use low-band, mid-band, and high-band spectrum, and 3GPP is in the process of identifying candidate bands for 5G across all frequency ranges. 3GPP Technical Services Group - Radio Access Networks (TSG-RAN) agreed to a process of efficiently adding LTE/NR band combinations and carrier-aggregated NR/NR band combinations. Initial bands being considered by the 3GPP specification process include low band bands 5, 20, 28 and 71, mid-band bands 1, 3, 7, 66 and 41, and key mmWave bands. Just as it has done with LTE, over time, 3GPP will specify additional 5G bands spanning multiple frequencies.

Although 5G research and development is in its infancy, to achieve the 20 Gbps or higher throughput rates envisioned for 5G will require radio carriers such as 200 MHz or 400 MHz, bandwidths available only at frequencies above 5 GHz. Researchers globally are studying high-frequency spectrum options including both cmWave frequencies (3 GHz to 30 GHz) and mmWave (30 GHz to 300 GHz). Ten times as much spectrum, or more, could be available in these higher frequencies than in all current cellular spectrum.

---

[83] For more details, see Sprint press release, "Qualcomm, SoftBank and Sprint Announce Collaboration on 2.5 GHz 5G," May 10, 2017, available at http://newsroom.sprint.com/qualcomm-softbank-and-sprint-announce-collaboration-on-25-ghz-5g.htm.

[84] For more details, see Sprint article at: http://newsroom.sprint.com/blogs/sprint-perspectives/introducing-the-sprint-lte-plus-network--faster-stronger-more-reliable-than-ever-before.htm

5G NR is being designed to operate in any frequency band. These bands encompass 600 MHz, current cellular bands, 3.5 GHz, and higher bands such as mmWave. Because 5G will be the first cellular technology to be able to operate in mmWave bands, industry and government are working to make such bands available for 5G.

During the World Radiocommunication Conference (WRC) 15, the ITU proposed a set of global frequencies for 5G[85], which it intends to finalize at the next conference in 2019 (WRC 19):

- ❑ 24.25–27.5GHz

- ❑ 31.8–33.4GHz

- ❑ 37–40.5GHz

- ❑ 40.5–42.5GHz

- ❑ 45.5–50.2GHz

- ❑ 50.4–52.6GHz

- ❑ 66–76GHz

- ❑ 81–86GHz

- ❑ In 2014, the FCC published a Notice of Inquiry into use of spectrum bands above 24 GHz for Mobile Radio Services,[86] followed by a Notice of Proposed Rulemaking in October 2015. [87]

The FCC issued adopted rules to identify and open up 5G mmWave spectrum allocation in July 2016 that identify 3.85 GHz of licensed spectrum and 7 GHz of unlicensed spectrum: licensed use in 28 GHz, 37 GHz, and 39 GHz bands; unlicensed use in 64-71 GHz; and shared access in the 37-37.6 GHz band.

On July 24, 2016, the FCC adopted a report and order and further notice of proposed rulemaking that creates a new Part 30 for rules governing 28 GHz, 36 GHz, and 39 GHz bands, summarized in Table 12.

Table 12: United States 5G mmWave Bands[88]

Bands	Details
28 GHz Band (27.5-28.35 GHz)	Currently licensed for Local Multipoint Distribution Service (LMDS). Could be licensed using either 200 or 400 MHz wide channels.

---

[85] 5G Americas Webcast, "LTE-Steps to 5G," Feb 12, 2016.

[86] FCC, *Notice of Inquiry, Use of Spectrum Bands above 24 GHz for Mobile Radio Services*, Oct. 17, 2014.

[87] FCC, *Notice of Proposed Rulemaking, Use of Spectrum Bands Above 24 GHz for Mobile Radio Services, GN Docket No. 14-177,* Oct 23, 2015.

[88] For more details, refer to FCC, *Report and Order and Further Notice of Proposed Rulemaking, Use of Spectrum Bands Above 24 GHz for Mobile Radio Services*, July 14, 2016. See also 5G Americas, *5G*

Bands	Details
39 GHz Band (38.6-40 GHz)	Currently licensed for fixed microwave in 50 MHz channels. Segment will be auctioned in 200 MHz blocks.
37 GHz Band (37-38.6 GHz)	Lower 37-37.6 GHz segment will be shared between federal and non-federal users. Upper 37.6-38.6 GHz segment will be auctioned in 200 MHz blocks.
64-71 GHz Band	Available for unlicensed use with same Part 15 rules as existing 57-64 GHz band.

In the 28 GHz band, satellite operations are secondary, but these operations are co-primary in the 37/39 GHz bands. Spectrum sharing may be required in some 5G bands, including 38.6 to 40 GHz, such as with fixed satellite service.

In the further notice of proposed rulemaking, the FCC is investigating additional bands, including 24 GHz, 32 GHz, 42 GHz, 47 GHz, 50 GHz, 70/80 GHz, and bands above 95 GHz.

The complex ITU harmonization process may mean that some regions, or even countries, pursue 5G bands that are not globally harmonized. For example, U.S. operators, along with operators in Korea and Japan, are planning 5G trials in the 28 GHz band, even though it is not one of the ITU bands.

Harmonization

Spectrum harmonization delivers many benefits, including higher economies of scale, better battery life, improved roaming, and reduced interference along borders.

As regulators make more spectrum available, it is important that they follow guidelines such as those espoused by 5G Americas: [89]

- ❑ Configure licenses with wider bandwidths.
- ❑ Group like services together.
- ❑ Be mindful of global technology standards.
- ❑ Pursue harmonized/contiguous spectrum allocations.
- ❑ Exhaust exclusive use options before pursuing shared use.
- ❑ Because not all spectrum is fungible, align allocation with demand.

Emerging technologies such as LTE benefit from wider radio channels. These wider channels are not only spectrally more efficient, they also offer greater capacity. Figure 34 shows increasing LTE spectral efficiency obtained with wider radio channels, with 20 MHz on the

---

Spectrum Recommendations, April 2017, available at http://www.5gamericas.org/files/9114/9324/1786/5GA_5G_Spectrum_Recommendations_2017_FINAL.pdf. This report covers global 5G spectrum developments.

[89] 4G Americas, *Sustaining the Mobile Miracle – A 4G Americas Blueprint for Securing Mobile Broadband Spectrum in this Decade*, March 2011.

downlink and 20 MHz (20+20 MHz) on the uplink comprising the most efficient configuration.
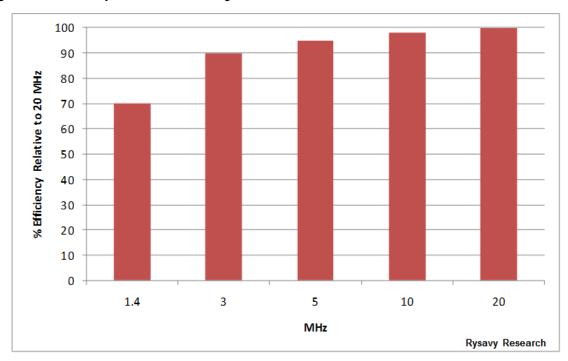
Figure 34: LTE Spectral Efficiency as Function of Radio Channel Size[90]

Rysavy Research

The organization tasked with global spectrum harmonization, the International Telecommunication Union, periodically holds World Radiocommunication Conferences.[91]

Harmonization occurs at multiple levels:

❑ Allocation of radio frequencies to a mobile service in the ITU frequency allocation table.

❑ Establishment of global or regional frequency arrangements, including channel blocks and specific duplexing modes.

❑ Development of detailed technical specifications and standards, including system performance, RF performance, and coexistence with other systems in neighboring bands.

❑ Assignment for frequency blocks with associated technical conditions and specifications to appropriate operators and service providers.[92]

---

[90] 5G Americas member company analysis.

[91] International Telecommunication Union, "World Radiocommunication Conferences (WRC)," http://www.itu.int/ITU-R/index.asp?category=conferences&rlink=wrc&lang=en, viewed May 18, 2017.

[92] International Telecommunication Union Radiocommunication Study Groups, *Technical Perspective on Benefits Of Spectrum Harmonization for Mobile Services and IMT*, Document 5D/246-E, January 2013.

Unlicensed Spectrum

Wi-Fi uses spectrum efficiently because its small coverage areas result in high-frequency reuse and high data density (bps per square meter). Less efficient are white-space unlicensed networks, sometimes called "super Wi-Fi," that, because of large coverage areas, have much lower throughput per square meter. While white-space networks may be a practical broadband solution in rural or undeveloped areas, they face significant challenges in urban areas that already have mobile and fixed broadband available.[93] See the section on "White Space Networks" in the appendix for further details.

Advocates argue that unlicensed spectrum unleashes innovation and that government should allocate greater amounts of unlicensed spectrum. Although Wi-Fi has been successful, the core elements that make unlicensed spectrum extremely successful are also the source of inherent disadvantages: local coverage and its unlicensed status. Local coverage enables high data density and high frequency reuse but makes widespread continuous coverage almost impossible. Similarly, unlicensed operation facilitates deployment by millions of entities but results in overlapping coverage and interference.

Networks built using unlicensed spectrum cannot replace networks built using licensed spectrum, and vice versa. The two are complementary and helpful to each other, as summarized in Table 13.[94]

Table 13: Pros and Cons of Unlicensed and Licensed Spectrum
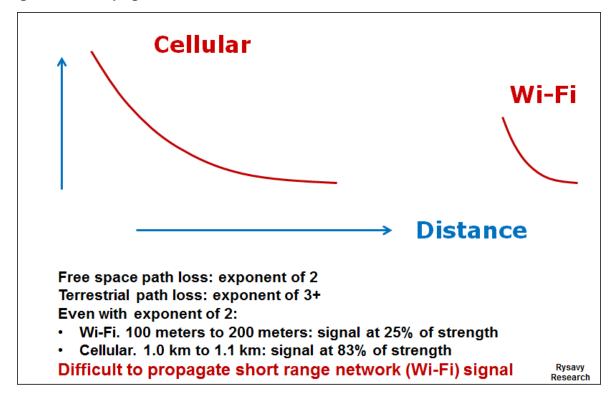
Unlicensed Spectrum		Licensed Spectrum	
Pros	Cons	Pros	Cons
Easy and quick to deploy	Potential of other entities using same frequencies	Huge coverage areas	Expensive infrastructure
Low-cost hardware	Difficult to impossible to provide wide-scale coverage	Able to manage quality of service	Each operator has access to only a small amount of spectrum

Some operators offer a "Wi-Fi first" capability with which devices always attempt to use a Wi-Fi connection and fall back to a cellular connection only if no Wi-Fi is available. Such cellular backup is essential because Wi-Fi, due to low-power operation in many bands, is inherently unsuited for providing continuous coverage. The sharp drop-off in signal strength due to low transmit power makes coverage gaps over large areas inevitable, especially outdoors.

---

[93] For further analysis, see Rysavy Research, "White spaces networks are not 'super' nor even Wi-Fi," Gigaom, March 2013. Available at http://gigaom.com/2013/03/17/white-spaces-networks-are-not-super-nor-even-wi-fi/.

[94] For further analysis, see Rysavy Research, "It's Time for a Rational Perspective on Wi-Fi," Gigaom, April, 2014. Available at http://gigaom.com/2014/04/27/its-time-for-a-rational-perspective-on-wi-fi/.

Figure 35: Propagation Losses of Cellular vs. Wi-Fi[95]

Free space path loss: exponent of 2
Terrestrial path loss: exponent of 3+
Even with exponent of 2:
- Wi-Fi. 100 meters to 200 meters: signal at 25% of strength
- Cellular. 1.0 km to 1.1 km: signal at 83% of strength

Difficult to propagate short range network (Wi-Fi) signal

Rysavy Research

Spectrum Sharing

In 2012, the President's Council of Advisors on Science and Technology issued a report titled, "Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth." The report recommended spectrum sharing between government and commercial entities.

The U.S. government can designate spectrum for exclusive, shared, or unlicensed use, as shown in Figure 36. Shared use can be opportunistic, as with TV white spaces; two-tier with incumbents and licensed users; or three-tier, which adds opportunistic access. The bands initially targeted for spectrum sharing include AWS-3 (two tiers on a temporary basis) and the 3.5 GHz band (three tiers).

The three-tier plan envisioned by the U.S. government for the 3.5 GHz band gives more entities access to the spectrum but at the cost of increased complexity.

---

[95] Assumes 1.0 km radius for cellular and 100-meter radius for Wi-Fi.

Figure 36: Spectrum Use and Sharing Approaches

The European Telecommunications Standards Institute (ETSI) is the leading organization standardizing cognitive radios. The most relevant effort is called "Licensed Shared Access" (LSA), a two-tier spectrum sharing system that includes incumbents and licensed secondary users that access shared spectrum via a database, as depicted in Figure 37.

Figure 37: Licensed Shared Access (LSA)

The three-tier system expected for the 3.5 GHz band in the United States will be complex, necessitating a real-time Spectrum Access System, the SAS, the design and development of which will encompass:

❑ Algorithms and methods;

❑ Methods of nesting hierarchical SAS entities (federal secure SAS versus commercial SASs);

- ❑ Coordination among multiple, competing commercial SAS managing entities;

- ❑ Interface definitions;

- ❑ Communication protocol definitions;

- ❑ Database and protocol security;

- ❑ Policy enforcement;

- ❑ Speed of channel allocation/reallocation;

- ❑ Time intervals for spectrum allocation;

- ❑ Effectively managing large numbers of Tier 3 users; and

- ❑ Data ownership, fees, rules, fairness, and conflict resolutions, all of which have policy, regulatory, and business implications.

TVWS databases available today address only a tiny subset of these requirements.[96]

Figure 38 shows the proposed architecture of the 3.5 GHz system. The system consists of incumbents (government systems), Priority Access Licenses, and General Authorized Access. Government systems include military ship-borne radar, military ground-based radar, fixed satellite service earth stations (receive only), and government broadband services (3650 to 3700 MHz). PAL licenses will be used by entities such as cellular operators and will be available for three-year periods. GAA users are licensed by rule and must protect both incumbents and PALs. Incumbents are protected by an Environmental Sensing Capability that detects incumbents and informs the SAS. Some examples of GAA use cases are small-business hotspots, campus hotspots, and backhaul.

Citizens Broadband Radio Service Devices (CBSDs) are the base stations operating under this service; they can operate only under the authority and management of the SAS, either by direct communications or a proxy node.

---

[96] For further analysis, see Rysavy Research and Mobile Future, *Complexities of Spectrum Sharing: How to Move Forward*, April 2014. Available at http://www.rysavy.com/Articles/2014-04-Spectrum-Sharing-Complexities.pdf.
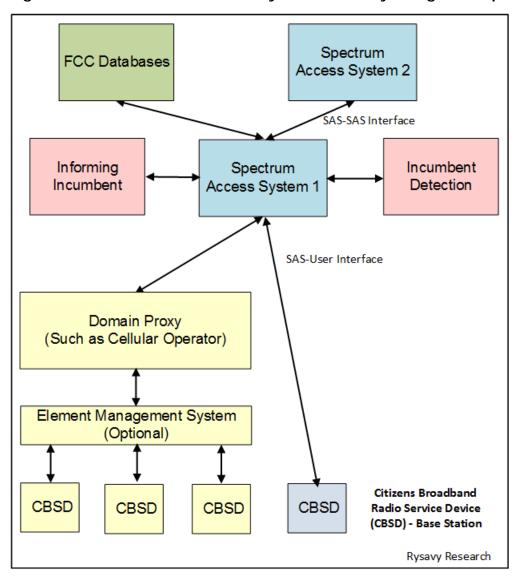
Figure 38: United States 3.5 GHz System Currently Being Developed

Conclusion

Mobile broadband remains at the forefront of innovation and development in computing, networking, and application development. As users, and now machines, consume ever more wireless data, the industry is responding with more efficient, faster, and higher capacity networks. Even as excitement builds about 5G, LTE, through ongoing advances, has become the global standard.

LTE-Advanced and LTE-Advanced Pro innovations include VoLTE, 1 Gbps peak rate capability, higher-order MIMO, carrier aggregation, LAA/LWA/LWIP, IoT capabilities in Narrowband-IoT and Category M-1, V2X communications, small-cell support, URLLC, SON, dual connectivity, and CoMP—all capabilities that will improve performance, efficiency, and capacity and enable support for new vertical segments. Carriers are implementing NFV and SDN to reduce network costs, improve service velocity, and simplify deployment of new services. Such improvements also facilitate cloud RANs that promise further efficiency gains.

5G research and development efforts have accelerated; standards-based deployment will begin in 2019 and continue through 2030. Operators have announced pre-standard fixed-wireless deployments for 2017. 3GPP has implemented a 5G standardization process that began in Release 14 with a study of the new 5G radio, continues now with a first phase of specifications in Release 15 that provides both non-standalone and standalone options, then moves ahead with a second phase of complete specifications in Release 16.

By harnessing new spectrum, such as mmWave bands above 24 GHz, 5G will eventually be able to access ten times as much spectrum as is currently available for cellular operation. Using radio bands 200 MHz or wider will result in multi-Gbps throughput capabilities. 5G will be designed to integrate with LTE networks, and many 5G features may be implemented as LTE-Advanced Pro extensions prior to full 5G availability.

Small cells will play an ever-more-important role in boosting capacity, benefiting from a number of technologies and developments, including SON, eICIC, Dual Connectivity, LWA/LWIP, LTE-U, LTE-LAA, MulteFire, improved backhaul options, and spectrum ideal for small cells, such as the 3.5 GHz and mmWave bands.

Obtaining more spectrum remains a priority globally. In U.S. markets, the recently completed television incentive auctions for 600 MHz spectrum supplied 70 MHz of new licensed spectrum; the 3.5 GHz small-cell band will provide 150 MHz; and the 5G spectrum bands, with both licensed and unlicensed bands, will add about 10 GHz of new spectrum.

The future of mobile broadband, including both LTE-Advanced and 5G, is bright, with no end in sight for continued growth in capability, nor for the limitless application innovation that mobile broadband enables.

Appendix: Technology Details

The 3GPP family of data technologies provides ever increasing capabilities that support ever more demanding applications. Services obviously need to provide broad coverage and high data throughput. Less obvious for users, but as critical for effective application performance, are the need for low latency, QoS control, and spectral efficiency. Higher spectral efficiency translates to higher average throughputs (and thus more responsive applications) for more active users in a coverage area. The discussion below details the progression of capability for each technology, including throughput, security, latency, QoS, and spectral efficiency.

This appendix provides details on 5G, UMTS/HSPA, and EDGE.

3GPP Releases

❑ **Release 99**: Completed. First deployable version of UMTS. Enhancements to GSM data (EDGE). Provides support for GSM/EDGE/GPRS/WCDMA radio-access networks.

❑ **Release 4**: Completed. Multimedia messaging support. First steps toward using IP transport in the core network.

❑ **Release 5**: Completed. HSDPA. First phase of Internet Protocol Multimedia Subsystem (IMS). Full ability to use IP-based transport instead of just Asynchronous Transfer Mode (ATM) in the core network.

❑ **Release 6**: Completed. HSUPA. Enhanced multimedia support through Multimedia Broadcast/Multicast Services (MBMS). Performance specifications for advanced receivers. Wireless Local Area Network (WLAN) integration option. IMS enhancements. Initial VoIP capability.

❑ **Release 7**: Completed. Provides enhanced GSM data functionality with Evolved EDGE. Specifies HSPA+, which includes higher-order modulation and MIMO. Performance enhancements, improved spectral efficiency, increased capacity, and better resistance to interference. Continuous Packet Connectivity (CPC) enables efficient "always-on" service and enhanced uplink UL VoIP capacity, as well as reductions in call set-up delay for Push-to-Talk Over Cellular (PoC). Radio enhancements to HSPA include 64 Quadrature Amplitude Modulation (QAM) in the downlink and 16 QAM in the uplink. Also includes optimization of MBMS capabilities through the multicast/broadcast, single frequency network (MBSFN) function.

❑ **Release 8**: Completed. Comprises further HSPA Evolution features such as simultaneous use of MIMO and 64 QAM. Includes dual-carrier HSDPA (DC-HSDPA) wherein two downlink carriers can be combined for a doubling of throughput performance. Specifies OFDMA-based 3GPP LTE. Defines EPC and EPS.

❑ **Release 9**: Completed. HSPA and LTE enhancements including HSPA dual-carrier downlink operation in combination with MIMO, Multimedia Broadcast Multicast Services (MBMS), HSDPA dual-band operation, HSPA dual-carrier uplink operation, EPC enhancements, femtocell support, support for regulatory features such as emergency user equipment positioning and Commercial Mobile Alert System (CMAS), and evolution of IMS architecture.

❑ **Release 10**: Completed. Specifies LTE-Advanced that meets the requirements set by ITU's IMT-Advanced project. Key features include carrier aggregation, multi-antenna enhancements such as enhanced downlink eight-branch MIMO and uplink MIMO, relays,

enhanced LTE Self-Organizing Network capability, Evolved Multimedia Broadcast Multicast Services (eMBMS), HetNet enhancements that include eICIC, Local IP Packet Access, and new frequency bands. For HSPA, includes quad-carrier operation and additional MIMO options. Also includes femtocell enhancements, optimizations for M2M communications, and local IP traffic offload.

❑ **Release 11**: Completed. For LTE, emphasizes Coordinated Multi Point (CoMP), carrier-aggregation enhancements, devices with interference cancellation, development of the Enhanced Physical Downlink Control Channel (EPDCCH), and further enhanced eICIC including devices with CRS (Cell-specific Reference Signal) interference cancellation. The release includes further DL and UL MIMO enhancements for LTE. For HSPA, provides eight-carrier on the downlink, uplink enhancements to improve latency, dual-antenna beamforming and MIMO, CELL Forward Access Channel (FACH) state enhancement for smartphone-type traffic, four-branch MIMO enhancements and transmissions for HSDPA, 64 QAM in the uplink, downlink multipoint transmission, and noncontiguous HSDPA carrier aggregation. Wi-Fi integration is promoted through S2a Mobility over GPRS Tunneling Protocol (SaMOG). An additional architectural element called "Machine-Type Communications Interworking Function" (MTC-IWF) will more flexibly support machine-to-machine communications.

❑ **Release 12**: Completed. Enhancements include improved small cells/HetNets for LTE, LTE multi-antenna/site technologies (including Active Antenna Systems), Dual Connectivity, 256 QAM modulation option, further CoMP/MIMO enhancements, enhancements for interworking with Wi-Fi, enhancements for MTC, SON, support for emergency and public safety, Minimization of Drive Tests (MDT), advanced receivers, device-to-device communication (also referred to as Proximity Services), group communication enablers in LTE, addition of Web Real Time Communication (WebRTC) to IMS, energy efficiency, more flexible carrier aggregation, dynamic adaptation of uplink-downlink ratios in TDD mode, further enhancements for HSPA+, small cells/HetNets, Scalable-UMTS, and FDD-TDD carrier aggregation.

❑ **Release 13**: Completed. LTE features include Active Antenna Systems (AAS) with support for as many as 16 antenna elements (full-dimension MIMO) and beamforming, Network-Assisted Interference Cancellation and Suppression (NAICS), radio-access network sharing, carrier aggregation supporting 32 component carriers,[97] carrier aggregation of up to four carriers on the downlink and two carriers on the uplink, LAA for operation in unlicensed bands, LTE Wi-Fi Aggregation including LWIP, RCLWI, isolated operation and mission-critical voice communications for public safety, application-specific congestion management, User-Plane Congestion Management, enhancement to WebRTC interoperability, architecture enhancement for dedicated core networks, enhancement to proximity-based services, Mission-Critical Push-to-Talk, group communications, CoMP enhancements, small cell enhancements, machine-type communications enhancements including NB-IoT and Extended Coverage GSM (EC-GSM), VoLTE enhancements, SON enhancements, shared network enhancements, indoor positioning based on WLAN access points, Bluetooth beacons and barometric pressure, and enhanced circuit-switched fallback. HSPA+ features include support for dual-band uplink carrier aggregation.

---

[97] This level of aggregation refers to signaling capabilities. The number of carriers that can be combined in an actual deployment is smaller and depends on RAN co-existence studies. Refer to the appendix section on "Carrier Aggregation" for additional details.

- ❑ **Release 14**: Expected completion June 2017. Contemplated features include uplink operation for LAA (enhanced LAA), full-dimension MIMO enhanced with up to 32 antenna elements, dual-connectivity of licensed and unlicensed carriers across non-collocated nodes, vehicle-to-vehicle and vehicle-to-infrastructure (V2X) communications built on Release 12 Proximity Services, shared LTE broadcast in which different operators broadcast the same content on the same frequency, non-IP operation for IoT, Downlink Multi-user Superposition Transmission (MUST), enhanced LWA, VoLTE enhancements, LWIP/LWA enhancements, eMBMS enhancements, NB-IoT enhancements, and LTE latency reduction.

- ❑ **Release 15**: Expected completion September 2018. Non-standalone (using LTE core network) option expected March 2018. Specifies phase 1 of 5G. NR radio, 4G-5G interworking, and MIMO/beamforming. Further LTE enhancements include ultra-reliable low-latency communications, NB-IoT enhancements, LAA enhancements, V2X enhancements, DL 1024 QAM, CoMP enhancements, AAS enhancements, and LTE/5G core network capability.

- ❑ **Release 16**: Expected completion end of 2019. Specifies phase 2 of 5G. Adds URLLC, spectrum sharing, unlicensed spectrum operation and integration, and multiple other enhancements. Further LTE enhancements.

Data Throughput Comparison

Data throughput is an important metric for quantifying network throughput performance. Unfortunately, the ways in which various organizations quote throughputs vary tremendously, often resulting in misleading claims. The intent of this paper is to realistically represent the capabilities of these technologies.

One method of representing a technology's throughput is what people call "peak throughput" or "peak network speed," which refers to the fastest possible transmission speed over the radio link and is generally based on the highest-order modulation available and the least amount of coding (error correction) overhead. Peak network speed is also usually quoted at layer 2 of the radio link. Because of protocol overhead, actual application throughput may be up to 10% lower than this layer-2 value.

Another method is to disclose throughputs actually measured in deployed networks with applications such as File Transfer Protocol (FTP) under favorable conditions, which assume light network load (as low as one active data user in the cell sector) and favorable signal propagation. This number is useful because it demonstrates the high-end, actual capability of the technology in current deployments, referred to in this paper as the "peak user rate." Average rates are lower than this peak rate and are difficult to predict because they depend on a multitude of operational and network factors. Except when the network is congested, however, the majority of users should experience throughput rates higher than one-half of the peak achievable rate.

Some operators, primarily in the United States, also quote typical throughput rates, which are based on throughput tests the operators have done across their operating networks and incorporate a higher level of network load. Although the operators do not disclose the precise methodologies they use to establish these figures, the values provide a good indication of what users can realistically expect.

Table 14 presents the technologies in terms of peak network throughput rates, peak user rates (under favorable conditions), and typical rates. It omits values that are not yet known, such as for future technologies.

The projected typical rates for HSPA+ and LTE show a wide range because these technologies exploit favorable radio conditions to achieve high throughput rates, but under poor radio conditions, throughput rates are lower.

Table 14: Throughput Performance of Different Wireless Technologies (Blue Indicates Theoretical Peak Rates, Green Typical)

	Downlink		Uplink	
	Peak Network Speed	Peak and/or Typical User Rate	Peak Network Speed	Peak and/or Typical User Rate
5G in mmWave, early versions[98]	5 Gbps	500 Mbps	2 Gbps	250 Mbps
5G in mmWave, later versions	50 Gbps	5 Gbps	25 Gbps	2 Gbps
LTE (2X2 MIMO, 10+10 MHz, DL 64 QAM, UL 16 QAM)	70 Mbps	6.5 to 26.3 Mbps[99]	35 Mbps[100]	6.0 to 13.0 Mbps
LTE-Advanced (2X2 or 4X4 MIMO, 20+20 MHz or 40+20 MHz with Carrier Aggregation [CA], DL 64 QAM, UL 16 QAM)	300 Mbps	N/A	71 Mbps[101]	N/A
LTE Advanced (4X4 MIMO, 60+20MHz, CA, 256 QAM DL, 64 QAM UL)	600 Mbps		150 Mbps	
LTE Advanced (4X4 MIMO, 80+20 MHz, CA, 256 QAM DL, 64 QAM UL)	> 1 Gbps		150 Mbps	

---

[98] Speculative values, Rysavy Research estimates. Assumes 200 MHz radio channel, 2:1 TDD. Throughput rates would double using 400 MHz.

[99] 5G Americas member company analysis for downlink and uplink. Assumes single user with 50% load in other sectors. AT&T and Verizon are quoting typical user rates of 5-12 Mbps on the downlink and 2-5 Mbps on the uplink for their networks. See additional LTE throughput information in the section below, "LTE Throughput."

[100] Assumes 64 QAM. Otherwise 22 Mbps with 16 QAM.

[101] Assumes 64 QAM. Otherwise 45 Mbps with 16 QAM.

	Downlink		Uplink	
	Peak Network Speed	Peak and/or Typical User Rate	Peak Network Speed	Peak and/or Typical User Rate
LTE Advanced (8X8 MIMO, 20+20 MHz, DL 64 QAM, UL 64 QAM)	1.2 Gbps	N/A	568 Mbps	N/A
LTE Advanced, 100 MHz + 100 MHz	3 Gbps		1.5 Gbps	
LTE Advanced 32 Carriers	>> 3 Gbps			
EDGE (type 2 MS)	473.6 Kbps	Not Applicable (N/A)	473.6 Kbps	N/A
EDGE (type 1 MS) (Practical Terminal)	236.8 Kbps	200 Kbps peak 160 to 200 Kbps typical[102]	236.8 Kbps	200 Kbps peak 80 to 160 Kbps typical[103]
HSDPA Initial Devices (2006)	1.8 Mbps	> 1 Mbps peak	384 Kbps	350 Kbps peak
HSDPA	14.4 Mbps	N/A	384 Kbps	N/A

---

[102] Assumes four-to-five downlink timeslot devices (each timeslot capable of 40 Kbps).

[103] Assumes two-to-four uplink timeslot devices (each timeslot capable of 40 Kbps).

	Downlink		Uplink	
	Peak Network Speed	Peak and/or Typical User Rate	Peak Network Speed	Peak and/or Typical User Rate
HSPA[104] Initial Implementation	7.2 Mbps	> 5 Mbps peak 700 Kbps to 1.7 Mbps typical[105]	2 Mbps	> 1.5 Mbps peak 500 Kbps to 1.2 Mbps typical
HSPA	14.4 Mbps	N/A	5.76 Mbps	N/A
HSPA+ (DL 64 QAM, UL 16 QAM, 5+5 MHz)	21.6 Mbps	1.9 Mbps to 8.8 Mbps typical[106]	11.5 Mbps	1 Mbps to 4 Mbps typical
HSPA+ (2X2 MIMO, DL 16 QAM, UL 16 QAM, 5+5 MHz)	28 Mbps	N/A	11.5 Mbps	N/A
HSPA+ (2X2 MIMO, DL 64 QAM, UL 16 QAM, 5+5 MHz)	42 Mbps	N/A	11.5 Mbps	N/A
HSPA+ (DL 64 QAM, UL 16 QAM, Dual Carrier, 10+5 MHz)	42 Mbps	Approximate doubling of 5+5 MHz rates - 3.8 to 17.6 Mbps.	11.5 Mbps	1 Mbps to 4 Mbps typical
HSPA+ (2X2 MIMO DL, DL 64 QAM, UL 16 QAM, Dual Carrier, 10+10 MHz)	84 Mbps	N/A	23 Mbps	N/A

[104] High Speed Packet Access (HSPA) consists of systems supporting both High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA).

[105] Typical downlink and uplink throughput rates based on AT&T press release, June 4, 2008

[106] 5G Americas member company analysis. Assumes Release 7 with 64 QAM and F-DPCH. Single user. 50% loading in neighboring cells. Higher rates expected with subsequent 3GPP releases.

	Downlink		Uplink	
	Peak Network Speed	**Peak and/or Typical User Rate**	**Peak Network Speed**	**Peak and/or Typical User Rate**
HSPA+ (2X2 MIMO DL, DL 64 QAM, UL 16 QAM, Quad-Carrier,[107] 20+10 MHz)	168 Mbps	N/A	23 Mbps	N/A
HSPA+ (2X2 MIMO DL and UL, DL 64 QAM, UL 16 QAM, Eight-Carrier, 40+10 MHz)	336 Mbps	N/A	69 Mbps	N/A
HSPA+ (4X2 MIMO DL, 2X2 MIMO UL, DL 64 QAM, UL 16 QAM, 8 carrier, 40+10 MHz)	672 Mbps	N/A	69 Mbps	N/A
EDGE (type 2 MS)	473.6 Kbps	Not Applicable (N/A)	473.6 Kbps	N/A
EDGE (type 1 MS) (Practical Terminal)	236.8 Kbps	200 Kbps peak 160 to 200 Kbps typical[108]	236.8 Kbps	200 Kbps peak 80 to 160 Kbps typical[109]
CDMA2000 EV-DO Rel. 0	2.4 Mbps	> 1 Mbps peak	153 Kbps	150 Kbps peak

---

[107] No operators have announced plans to deploy HSPA in a quad (or greater) carrier configuration. Three carrier configurations, however, have been deployed.

[108] Assumes four-to-five downlink timeslot devices (each timeslot capable of 40 Kbps).

[109] Assumes two-to-four uplink timeslot devices (each timeslot capable of 40 Kbps).

	Downlink		Uplink	
	Peak Network Speed	**Peak and/or Typical User Rate**	**Peak Network Speed**	**Peak and/or Typical User Rate**
CDMA2000 EV-DO Rev. A	3.1 Mbps	> 1.5 Mbps peak 600 Kbps to 1.4 Mbps typical[110]	1.8 Mbps	> 1 Mbps peak 300 to 500 Kbps typical
CDMA2000 EV-DO Rev. B (3 radio channels 5+5 MHz)	14.7[111] Mbps	Proportional increase of Rev A typical rates based on number of carriers.	5.4 Mbps	N/A
CDMA2000 EV-DO Rev B Theoretical (15 radio channels 20+20 MHz)	73.5 Mbps	N/A	27 Mbps	N/A

Additional information about LTE throughput appears below in the section "LTE Throughput."

Latency Comparison

As important as throughput is network latency, defined as the round-trip time it takes data to traverse the network. Each successive data technology from GPRS forward reduces latency, with LTE networks having latency as low as 15 msec. Ongoing improvements in each technology mean that all of these values will go down as vendors and operators fine-tune their systems. Figure 39 shows the latency of different 3GPP technologies.

---

[110] Typical downlink and uplink throughput rates based on Sprint press release Jan. 30, 2007.

[111] Assuming use of 64 QAM.
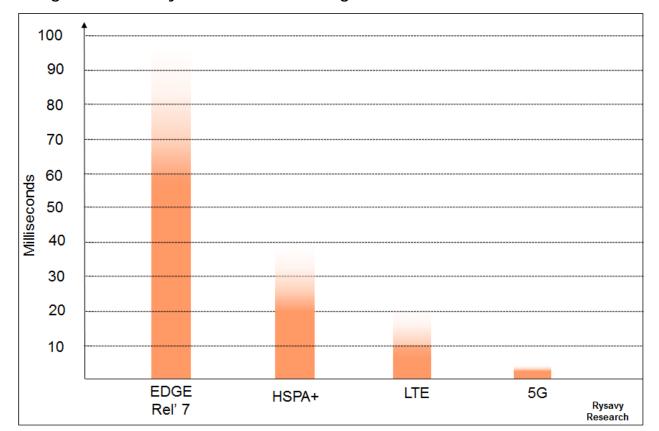
Figure 39: Latency of Different Technologies[112]

The values shown in Figure 39 reflect measurements of commercially deployed technologies, with EDGE Release 7 achieving 70 to 95 msec, HSPA+ 25 to 30 msec, and LTE 15 to 20 msec. A latency goal for 5G is less than 4 msec for broadband and 0.5 msec for mission-critical applications.

Spectral Efficiency

The evolution of data services is characterized by an increasing number of users with ever-higher bandwidth demands. As the wireless data market grows, deploying wireless technologies with high spectral efficiency is of paramount importance. Keeping all other things equal, including frequency band, amount of spectrum, and cell site spacing, an increase in spectral efficiency translates to a proportional increase in the number of users supported at the same load per user—or, for the same number of users, an increase in throughput available to each user.
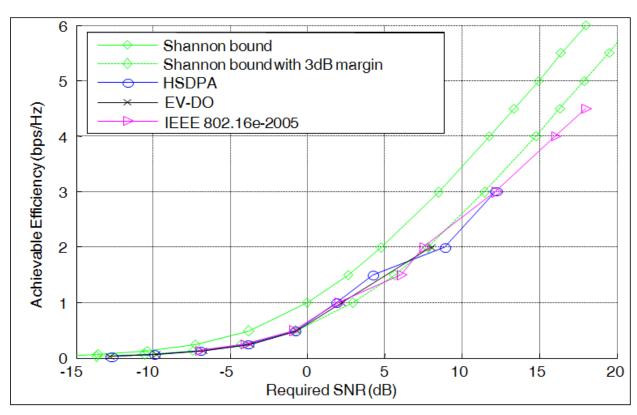
Increased spectral efficiency, however, comes at a price because it generally involves greater complexity for both user and base station equipment. Complexity can arise from the increased number of calculations performed to process signals or from additional radio

---

[112] 5G Americas member companies. Measured between subscriber unit and a node immediately external to wireless network. Does not include internet latency. Note that there is some variation in latency based on network configuration and operating conditions.

components. Hence, operators and vendors must balance market needs against network and equipment costs. OFDMA technologies, such as LTE and planned 5G approaches, achieve higher spectral efficiency with lower overall complexity, especially in larger bandwidths.

As shown in Figure 40, the link-layer performance of modern wireless technologies is approaching the theoretical limits as defined by the Shannon bound. (The Shannon bound is a theoretical limit to the information transfer rate [per unit bandwidth] that can be supported by any communications link. The bound is a function of the SNRs of the communications link.) Figure 40 also shows that HSDPA, 1xEV-DO, and IEEE 802.16e-2005 are all within 2 to 3 decibels (dB) of the Shannon bound, indicating that there is not much room for improvement from a link-layer perspective.

Figure 40: Performance Relative to Theoretical Limits for HSDPA, EV-DO, and WiMAX (IEEE 802.16e-2005)[113]

The curves in Figure 40 are for an Additive White Gaussian Noise Channel (AWGN). If the channel is slowly varying and the frame interval is significantly shorter than the coherence time, the effects of fading can be compensated for by practical channel estimation algorithms—thus justifying the AWGN assumption. For instance, at 3 km per hour and fading at 2 GHz, the Doppler spread is about 5.5 Hz. The coherence time of the channel is thus 1 second (sec)/5.5 or 180 msec. Frames are well within the coherence time of the channel, because they are typically 20 msec or less. As such, the channel appears

---

[113] 5G Americas member contribution.

"constant" over a frame, and the Shannon bound applies. Furthermore, significantly more of the traffic in a cellular system is at slow speeds (for example, 3 km/hr. or less) rather than at higher speeds. The Shannon bound is consequently also relevant for a realistic deployment environment.

As the speed of the mobile station increases and the channel estimation becomes less accurate, additional margin is needed. This additional margin, however, would impact the different standards fairly equally.

The focus of future technology enhancements is on improving system performance aspects that reduce interference to maximize the experienced SNRs in the system and antenna techniques (such as MIMO) that exploit multiple links or steer the beam rather than on investigating new air interfaces that attempt to improve link-layer performance.

MIMO techniques using spatial multiplexing to increase the overall information transfer rate by a factor proportional to the number of transmit or receive antennas do not violate the Shannon bound because the per-antenna transfer rate (that is, the per-communications link transfer rate) is still limited by the Shannon bound.

Figure 41 compares the spectral efficiency of different wireless technologies based on a consensus view of 5G Americas contributors to this paper. It shows the continuing evolution of the capabilities of all the technologies discussed. The values shown are reasonably representative of real-world conditions. Most simulation results produce values under idealized conditions; as such, some of the values shown are lower (for all technologies) than the values indicated in other papers and publications. For instance, 3GPP studies indicate higher HSPA and LTE spectral efficiencies. Nevertheless, there are practical considerations in implementing technologies that can prevent actual deployments from reaching calculated values. Consequently, initial versions of technology may operate at lower levels but then improve over time as designs are optimized. Therefore, readers should interpret the values shown as achievable, but not as the actual values that might be measured in any specific deployed network.
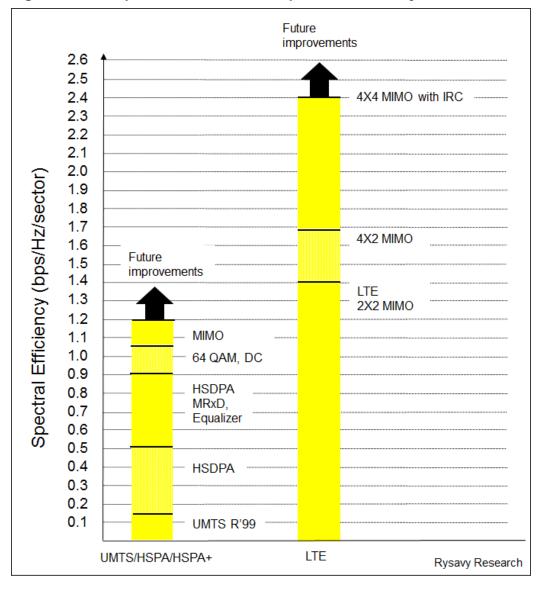
Figure 41: Comparison of Downlink Spectral Efficiency[114]

The values shown in Figure 41 are not all possible combinations of available features. Rather, they are representative milestones in ongoing improvements in spectral efficiency. For instance, terminals may employ Mobile Receive Diversity but not equalization.

The figure does not include EDGE, but EDGE itself is spectrally efficient at 0.6 bps/Hz using mobile receive diversity and, potentially, 0.7 bps/Hz with MIMO. Relative to WCDMA Release 99, HSDPA increases capacity by almost a factor of three. Type 3 receivers that include MMSE equalization and Mobile Receive Diversity (MRxD) effectively double HSDPA spectral efficiency. The addition of dual-carrier operation and 64 QAM increases spectral efficiency by about 15%, and MIMO can increase spectral efficiency by another 15%,

---

[114] Joint analysis by 5G Americas members. 5+5 MHz FDD for UMTS-HSPA/LTE. Mix of mobile and stationary users.

reaching 1.2 bps/Hz. Dual-carrier HSPA+ offers a gain in spectral efficiency from cross-carrier scheduling with possible gains of about 10%.[115]

Some enhancements, such as 64 QAM for HSPA, are simpler to deploy than other enhancements, such as 2X2 MIMO. The former can be done as a software upgrade, whereas the latter requires additional hardware at the base station. Thus, the figure does not necessarily show the actual progression of technologies that operators will deploy to increase spectral efficiency.

Beyond HSPA, 3GPP LTE results in further spectral efficiency gains, initially with 2X2 MIMO, then 4X2 MIMO, and then 4X4 MIMO. The gain for 4X2 MIMO will be 20% more than LTE with 2X2 MIMO; the gain for 4X4 MIMO in combination with interference rejection combining (IRC) will be 70% greater than 2X2 MIMO, reaching 2.4 bps/Hz. This value represents a practical deployment of 4X4 MIMO, with random phase and some timing-alignment error included in each of the four transmit paths. CoMP, discussed below in the appendix, provides a minimal contribution to spectral efficiency.

LTE is even more spectrally efficient when deployed using wider radio channels of 10+10 MHz and 20+20 MHz, although most of the gain is realized at 10+10 MHz. LTE TDD has spectral efficiency that is within 1% or 2% of LTE FDD.[116]

Figure 42 compares the uplink spectral efficiency of the different systems.

---

[115] 5G Americas member analysis. Vendor estimates for spectral-efficiency gains from dual-carrier operation range from 5% to 20%. Lower spectral efficiency gains are due to full-buffer traffic assumptions. In more realistic operating scenarios, gains will be significantly higher.

[116] Assumes best-effort traffic. Performance between LTE-TDD and FDD differs for real-time traffic for the following reasons: a.) The maximum number of HARQ process should be made as small as possible to reduce the packet re-transmission latency. b.) In FDD, the maximum number of HARQ process is fixed and, as such, the re-transmission latency is 7ms. c.) For TDD, the maximum number of HARQ process depends on the DL:UL configurations. As an example, the re-transmission latency for TDD config-1 is 9ms. d.) Because of higher re-transmission latency, the capacity of real-time services cannot be scaled for TDD from FDD based on the DL:UL ratio.
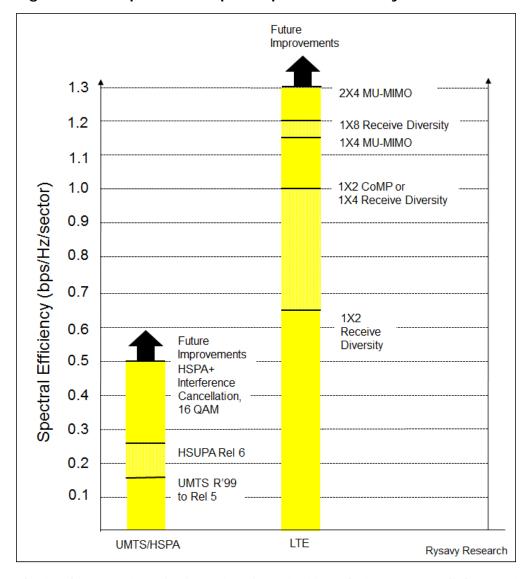
Figure 42: Comparison of Uplink Spectral Efficiency[117]

The implementation of HSUPA in HSPA significantly increases uplink capacity.

With LTE, spectral efficiency increases by use of receive diversity. Initial systems will employ 1X2 receive diversity (two antennas at the base station). 1X4 diversity will increase spectral efficiency by 50%, to 1.0 bps/Hz, and 1X8 diversity will provide a further 20% increase, from 1.0 bps/Hz to 1.2 bps/Hz.

It is also possible to employ Multi-User MIMO (MU-MIMO), which allows simultaneous transmission by multiple users on the same physical uplink resource to increase spectral efficiency. MU-MIMO will provide a 15% to 20% spectral efficiency gain, with actual increases depending on how well link adaptation is implemented. The figure uses a conservative 15% gain, showing MU-MIMO with a 1X4 antenna configuration increasing

---

[117] Joint analysis by 5G Americas members. 5+5 MHz for UMTS-HSPA/LTE. Mix of mobile and stationary users.

spectral efficiency by 15%, to 1.15 bps/Hz, and 2X4 MU-MIMO a further 15%, to 1.3 bps/Hz.

In Release 11, uplink CoMP using 1X2 increases efficiency from .65 bps/Hz to 1.0 bps/Hz. Many of the techniques used to improve LTE spectral efficiency can also be applied to HSPA since they are independent of the radio interface.

Figure 43 compares voice spectral efficiency.

Figure 43: Comparison of Voice Spectral Efficiency[118]
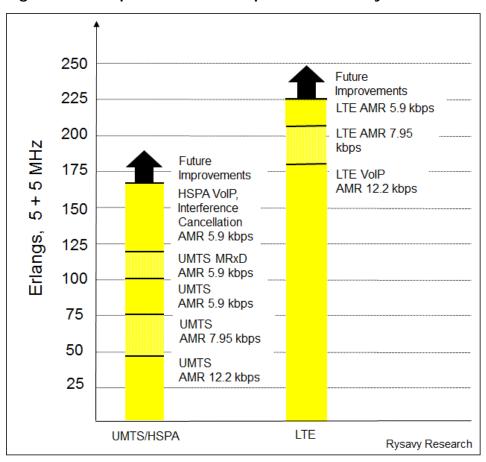
Figure 43 shows UMTS Release 99 with AMR 12.2 Kbps, 7.95 Kbps, and 5.9 Kbps vocoders. The AMR 12.2 Kbps vocoder provides superior voice quality in good (for example, static and indoor) channel conditions.

UMTS has dynamic adaptation between vocoder rates, enabling enhanced voice quality compared with EVRC at the expense of capacity in situations that are not capacity limited. With the addition of mobile receive diversity, UMTS circuit-switched voice capacity could reach 120 Erlangs in 5+5 MHz.

---

[118] Joint analysis by 5G Americas members. 5 + 5 MHz for UMTS-HSPA/LTE. Mix of mobile and stationary users.

VoIP Erlangs in this paper are defined as the average number of concurrent VoIP users that can be supported over a defined period of time (often one hour) assuming a Poisson arrival process and meeting a specified outage criteria (often less than 2% of the users exhibiting greater than 1% frame-error rate). Depending on the specific enhancements implemented, voice capacity could double over existing circuit-switched systems. These gains do not derive through use of VoIP, but rather from advances in radio techniques applied to the data channels. Many of these same advances may also be applied to current circuit-switched modes.

LTE achieves very high voice spectral efficiency because of better uplink performance since there is no in-cell interference. The figure shows LTE VoIP spectral efficiency using AMR at 12.2 Kbps, 7.95 Kbps, and 5.9 Kbps.

VoIP for LTE can use a variety of codecs. The figures show performance assuming specific codecs at representative bit rates. For Enhanced Variable Rate Codecs (EVRCs), the figure shows the average bit rate.

The voice efficiency of the wideband AMR voice codec, operating at 12.65 Kbps, is similar to the AMR codec at 12.2 Kbps, with a value of 180 Erlangs for both since both codecs operate at approximately the same bit rate. 1xRTT has voice capacity of 85 Erlangs in 5+5 MHz with EVRC-A and reaches voice capacity of 120 Erlangs with the use of Quasi-Linear Interference Cancellation (QLIC) and EVRC-B at 6 Kbps.

Data Consumed by Streaming and Virtual Reality

Table 15 quantifies usage based on advanced video compression schemes such as H.264 and H.265, the type of application, and usage per day.

Table 15: Data Consumed by Streaming and Virtual Reality[119]

Application	Throughput (Mbps)	MByte/hour	Hrs./day	GB/month
Audio or Music	0.1	58	0.5	0.9
			1.0	1.7
			2.0	3.5
			4.0	6.9
Small Screen Video (e.g., Feature Phone)	0.2	90	0.5	1.4
			1.0	2.7
			2.0	5.4
			4.0	10.8
Medium Screen Video (e.g., Smartphone, Tablet, Laptop)	1.0	450	0.5	6.8
			1.0	13.5
			2.0	27.0
			4.0	54.0
Larger Screen Video (e.g., 720p medium definition)	3.0	1350	0.5	20.3
			1.0	40.5
			2.0	81.0
			4.0	162.0
High Definition (e.g., 1080p Netflix HD)	5.0	2250	0.5	33.8
			1.0	67.5
			2.0	135
			4.0	270
4K Ultra-High Definition (Rates will range 12 to 30 Mbps)	20.0	9000	0.5	135
			1.0	270
			2.0	540
			4.0	1080
4G, 30 FPS, Virtual Reality (Rates will range 10 to 50 Mbps)	25.0	11250	0.5	169
			1.0	338
			2.0	675
			4.0	1350
8K, 90 FPS, Virtual Reality (Rates will exceed 200 Mbps)	200.0	90000	0.5	1350
			1.0	2700
			2.0	5400
			4.0	10800
6 Degrees Freedom VR (Rates will range 200 to 1,000 Mbps)	500.0	225000	0.5	3375
			1.0	6750
			2.0	13500
Rysavy Research			4.0	27000

---

[119] Rysavy Research analysis. For virtual reality-data requirements, refer to ABI Research/Qualcomm, *Augmented and Virtual Reality: the First Wave of 5G Killer Apps,* 2017. See also Netflix discussion of usage, "How can I control how much data Netflix uses?" https://help.netflix.com/en/node/87. Viewed May 3, 2016.

Spectrum Bands

3GPP technologies operate in a wide range of radio bands. As new spectrum becomes available, 3GPP updates its specifications for these bands. Although the support of a new frequency band may be introduced in a particular release, 3GPP specifies ways to implement devices and infrastructure operating on any frequency band, according to releases previous to the introduction of that particular frequency band. For example, although band 5 (US Cellular Band) was introduced in Release 6, the first devices operating on this band were compliant with the release 5 of the standard.

Table 16 shows the UMTS FDD bands.

Table 16: UMTS FDD Bands[120]

Operating Band	UL Frequencies UE Transmit, Node B Receive	DL Frequencies UE Receive, Node B Transmit
I	1920 - 1980 MHz	2110 -2170 MHz
II	1850 -1910 MHz	1930 -1990 MHz
III	1710-1785 MHz	1805-1880 MHz
IV	1710-1755 MHz	2110-2155 MHz
V	824 - 849MHz	869-894MHz
VI	830-840 MHz	875-885 MHz
VII	2500 - 2570 MHz	2620 - 2690 MHz
VIII	880 - 915 MHz	925 - 960 MHz
IX	1749.9 - 1784.9 MHz	1844.9 - 1879.9 MHz
X	1710-1770 MHz	2110-2170 MHz
XI	1427.9 - 1447.9 MHz	1475.9 - 1495.9 MHz
XII	699 - 716 MHz	729 - 746 MHz
XIII	777 - 787 MHz	746 - 756 MHz
XIV	788 - 798 MHz	758 - 768 MHz
XV	Reserved	Reserved
XVI	Reserved	Reserved
XVII	Reserved	Reserved
XVIII	Reserved	Reserved
XIX	830 – 845 MHz	875 -890 MHz
XX	832 - 862 MHz	791 - 821 MHz
XXI	1447.9 - 1462.9 MHz	1495.9 - 1510.9 MHz
XXII	3410 – 3490 MHz	3510 – 3590 MHz
XXV	1850 -1915 MHz	1930 -1995 MHz
XXVI	814-849 MHz	859-894 MHz
XXXII (NOTE 1)	N/A	1452 – 1496 MHz
NOTE 1: Restricted to UTRA operation when dual-band is configured (e.g., DB-DC-HSDPA or dual band 4C-HSDPA). The down link frequenc(ies) of this band are paired with the uplink frequenc(ies) of the other FDD band (external) of the dual band configuration.		

---

[120] 3GPP, *Base Station (BS) radio transmission and reception (FDD) (Release 14)*, December 2016, Technical Specification 25.104, V14.1.0.

Universal Mobile Telecommunications System (UMTS) Time Division Duplex (TDD) bands are the same as the LTE TDD bands. Table 17 details the LTE Frequency Division Duplex (FDD) and TDD bands.

Table 17: LTE FDD and TDD bands[121]

E-UTRA Operating Band	Uplink (UL) operating band BS receive UE transmit $F_{UL\_low} - F_{UL\_high}$		Downlink (DL) operating band BS transmit UE receive $F_{DL\_low} - F_{DL\_high}$		Duplex Mode
1	1920 MHz	– 1980 MHz	2110 MHz	– 2170 MHz	FDD
2	1850 MHz	– 1910 MHz	1930 MHz	– 1990 MHz	FDD
3	1710 MHz	– 1785 MHz	1805 MHz	– 1880 MHz	FDD
4	1710 MHz	– 1755 MHz	2110 MHz	– 2155 MHz	FDD
5	824 MHz	– 849 MHz	869 MHz	– 894MHz	FDD
6 (NOTE 1)	830 MHz	– 840 MHz	875 MHz	– 885 MHz	FDD
7	2500 MHz	– 2570 MHz	2620 MHz	– 2690 MHz	FDD
8	880 MHz	– 915 MHz	925 MHz	– 960 MHz	FDD
9	1749.9 MHz	– 1784.9 MHz	1844.9 MHz	– 1879.9 MHz	FDD
10	1710 MHz	– 1770 MHz	2110 MHz	– 2170 MHz	FDD
11	1427.9 MHz	– 1447.9 MHz	1475.9 MHz	– 1495.9 MHz	FDD
12	699 MHz	– 716 MHz	729 MHz	– 746 MHz	FDD
13	777 MHz	– 787 MHz	746 MHz	– 756 MHz	FDD
14	788 MHz	– 798 MHz	758 MHz	– 768 MHz	FDD
15	Reserved		Reserved		FDD
16	Reserved		Reserved		FDD
17	704 MHz	– 716 MHz	734 MHz	– 746 MHz	FDD
18	815 MHz	– 830 MHz	860 MHz	– 875 MHz	FDD
19	830 MHz	– 845 MHz	875 MHz	– 890 MHz	FDD
20	832 MHz	– 862 MHz	791 MHz	– 821 MHz	
21	1447.9 MHz	– 1462.9 MHz	1495.9 MHz	– 1510.9 MHz	FDD
22	3410 MHz	– 3490 MHz	3510 MHz	– 3590 MHz	FDD
23†	2000 MHz	– 2020 MHz	2180 MHz	– 2200 MHz	FDD
24	1626.5 MHz	– 1660.5 MHz	1525 MHz	– 1559 MHz	FDD
25	1850 MHz	– 1915 MHz	1930 MHz	– 1995 MHz	FDD
26	814 MHz	– 849 MHz	859 MHz	– 894 MHz	FDD
27	807 MHz	– 824 MHz	852 MHz	– 869 MHz	FDD
28	703 MHz	– 748 MHz	758 MHz	– 803 MHz	FDD
29	N/A		717 MHz	– 728 MHz	FDD (NOTE 2)
30	2305 MHz	– 2315 MHz	2350 MHz	– 2360 MHz	FDD
31	452.5 MHz	– 457.5 MHz	462.5 MHz	– 467.5 MHz	FDD
32	N/A		1452 MHz	– 1496 MHz	FDD (NOTE 2)
33	1900 MHz	– 1920 MHz	1900 MHz	– 1920 MHz	TDD
34	2010 MHz	– 2025 MHz	2010 MHz	– 2025 MHz	TDD
35	1850 MHz	– 1910 MHz	1850 MHz	– 1910 MHz	TDD
36	1930 MHz	– 1990 MHz	1930 MHz	– 1990 MHz	TDD
37	1910 MHz	– 1930 MHz	1910 MHz	– 1930 MHz	TDD
38	2570 MHz	– 2620 MHz	2570 MHz	– 2620 MHz	TDD
39	1880 MHz	– 1920 MHz	1880 MHz	– 1920 MHz	TDD
40	2300 MHz	– 2400 MHz	2300 MHz	– 2400 MHz	TDD
41	2496 MHz	– 2690 MHz	2496 MHz	– 2690 MHz	TDD
42	3400 MHz	– 3600 MHz	3400 MHz	– 3600 MHz	TDD
43	3600 MHz	– 3800 MHz	3600 MHz	– 3800 MHz	TDD
44	703 MHz	– 803 MHz	703 MHz	– 803 MHz	TDD
45	1447 MHz	– 1467 MHz	1447 MHz	– 1467 MHz	TDD
46	5150 MHz	– 5925 MHz	5150 MHz	– 5925 MHz	TDD (NOTE 3, NOTE 4)
47	5855 MHz	– 5925 MHz	5855 MHz	– 5925 MHz	TDD (NOTE 7)
48	3550 MHz	– 3700 MHz	3550 MHz	– 3700 MHz	TDD
65	1920 MHz	– 2010 MHz	2110 MHz	– 2200 MHz	FDD
66	1710 MHz	– 1780 MHz	2110 MHz	– 2200 MHz	FDD (NOTE 5)
67	N/A		738 MHz	– 758 MHz	FDD (NOTE 2)
68	698 MHz	– 728 MHz	753 MHz	– 783 MHz	FDD
69	N/A		2570 MHz	– 2620 MHz	FDD (NOTE 2)
70	1695 MHz	– 1710 MHz	1995 MHz	– 2020 MHz	FDD[6]

NOTE 1: Band 6, 23 are not applicable.
NOTE 2: Restricted to E-UTRA operation when carrier aggregation is configured. The downlink operating band is paired with the uplink operating band (external) of the carrier aggregation configuration that is supporting the configured Pcell.
NOTE 3: This band is an unlicensed band restricted to licensed-assisted operation using Frame Structure Type 3.
NOTE 4: Band 46 is divided into four sub-bands as in Table 5.5-1A.
NOTE 5: The range 2180 – 2200 MHz of the DL operating band is restricted to E-UTRA operation when carrier aggregation is configured.
NOTE 6: The range 2010-2020 MHz of the DL operating band is restricted to E-UTRA operation when carrier aggregation is configured and TX-RX separation is 300 MHz. The range 2005-2020 MHz of the DL operating band is restricted to E-UTRA operation when carrier aggregation is configured and TX-RX separation is 295 MHz.
NOTE 7: No BS requirement is defined for this band..

5G

This section provides early details on aspects of 5G, including architecture, LTE-NR coexistence, integrated access and backhaul, and performance.

Architecture

The overall 5G architecture consists of what 3GPP calls the New Generation Radio-Access Network (NG-RAN) and the 5G Core (5GC), as shown in Figure 44. The figure shows the Access and Mobility Management Function (AMF); the User Plane Function (UPF); the NR NodeB (gNB), which is the 5G base station; and the NG and Xn interfaces.

Figure 44: 5G Architecture[122]
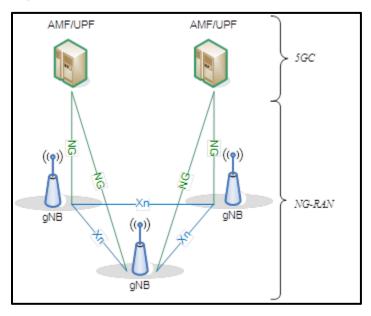
Figure 45 shows the functional split between the NG-RAN and 5GC.

---

[121] 3GPP, *Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception (Release 14)*, March 2017, Technical Specification 36.104, V14.3.0.

[122] 3GPP, *3GPP TS 38.300, NR; NR and NG-RAN Overall Description; Stage 2 (Release 15), V0.3.0 (2017-05).*

Figure 45: Functional Split between NG-RAN and 5GC[123]

The main body of this paper summarizes the features being specified in Releases 15 and 16 for NR and the core network. Additional capabilities that will be part of Release 15 include:

❑ A PDCP packet duplication function to allow redundant transmission of signaling or user data on two bearer paths.

❑ A new protocol layer called Service Data Adaptation Protocol (SDAP) that offers 5GC QoS flows.

❑ A new Radio Resource Control (RRC) inactive state designed for low-latency communications.

❑ A new system information broadcast model that allows on-demand system information instead of always having to broadcast system information (to reduce overhead in 5G beam sweeping).

Architecture Options

This topic was introduced in the main part of the paper and is covered here in more detail. In Release 15, 3GPP defines a number of different architecture options, shown in the following three figures. In many of these options, although not all, the 5G network integrates with LTE.

---

[123] Ibid.

Figure 46: 5G Network Architecture Options in 3GPP Release 15[124]
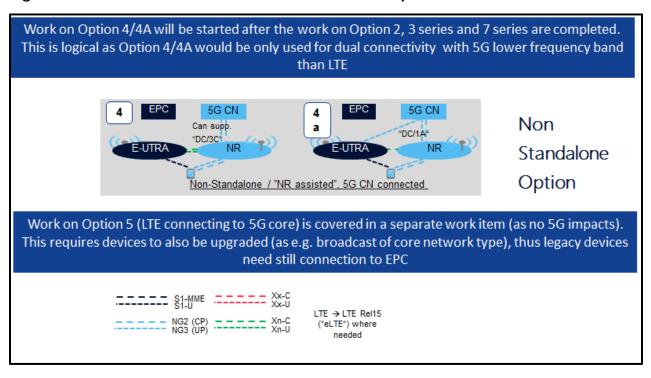
Figure 47: De-Prioritized 5G Network Architecture Options in 3GPP Release 15

---

[124] Nokia contribution, including subsequent two figures. For further details, refer to section 7.2, "5G Architecture Options," 3GPP TR 38.801, "Radio access architecture and interfaces."
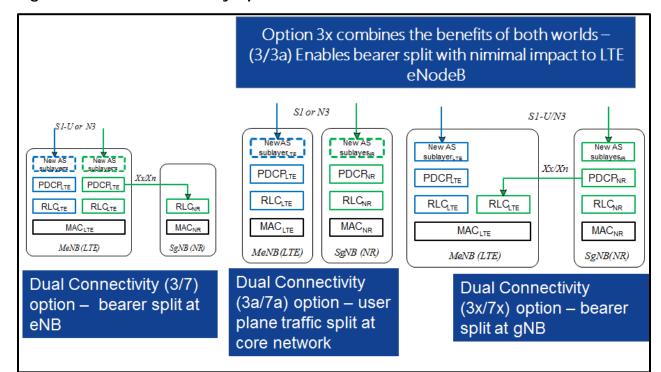
Figure 48: Dual-Connectivity Options with LTE as Master

LTE-NR Coexistence

LTE-NR coexistence is a Release 15 work item. This section describes how such co-existence may be achieved. Different LTE-NR co-existence cases include the following: time domain LTE/NR adjacent channel coexistence; LTE Secondary Cell on/off for LTE/NR adaptation; in-carrier LTE+NR coexistence in downlink, and in-carrier LTE+NR coexistence in uplink.

NR coexistence is required for LTE UEs of all releases. Because carrier aggregation was not introduced into LTE until LTE Release 10, CA-based techniques cannot be used as the sole means to achieve LTE/NR coexistence. However, CA techniques can be used for both time domain coexistence and frequency domain coexistence. For time domain coexistence, on a given carrier, LTE and NR are time-multiplexed by means of Secondary Cell (SCell) activation or deactivation. For frequency domain coexistence, the network configures a carrier, such as a 20MHz carrier, into multiple carriers, with, for example, 10MHz allocated to LTE and the remaining 10MHz to NR. Note that frequency domain coexistence can be accomplished without using carrier aggregation.

Figure 49 illustrates the frequency domain technique. Note that when splitting the 20MHz carrier into two allocations of 10MHz, the LTE carrier remains centered at the same frequency and the NR allocation is not consecutive.
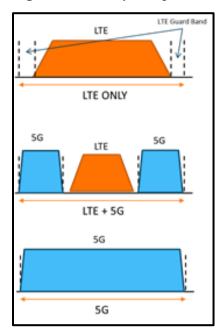
Figure 49: Frequency Domain Coexistence of LTE and NR[125]

Time Domain LTE/NR Coexistence Techniques

Time domain coexistence of LTE and NR can be dynamic (subframe level) or semi-static (MAC/RRC). In the latter case, spectrum resources are configured as SCell for a LTE UE, and the network can turn these resources on or off by means of SCell activation or deactivation using MAC control elements or by adding and removing the SCell via RRC signaling. Whenever the SCell is deactivated or not configured, the spectrum resources can be used for NR transmissions. In LTE Rel. 12, small-cell enhancements were introduced that allow a UE to measure discovery reference signals (DRS) on a deactivated SCell. In that case, further coordination between LTE and NR may be required even when the SCell is deactivated, as DRS may still be transmitted periodically. Generally, though, this kind of coexistence can be achieved by network implementation.

For the case of dynamic coexistence, LTE and NR co-exist in the same spectrum, and the network can multiplex the two on a subframe level. Because LTE transmits Cell-Specific Reference Signals (CRS) in all DL subframes and in the Downlink Pilot Time Slot (DwPTS) and non-Multicast-Broadcast Single-Frequency network (non-MBSFN) region of special and MBSFN subframes, respectively, dynamic coexistence is not as straightforward as semi-static coexistence.

Similar to the case of time domain coexistence based on CA techniques, whenever OFDM symbols do not carry CRS, coexistence can be achieved by a gNB scheduler implementation. In particular, the gNB can schedule mini-slot-based transmissions in the Uplink Pilot Time Slot (UpPTS) region of a special subframe and in the MBSFN region of an MBSFN subframe, neither of which carry CRS. In LTE UL subframes, the gNB can schedule NR transmissions using either slots or mini-slots. For example, when Sounding Reference Signal (SRS) is transmitted at the end of a subframe, all 14 OFDM symbols may not be available for NR, and mini-slots can be used. Otherwise, slots can be used to transmit NR

---

[125] AT&T contribution, including explanatory text.

signals and channels in LTE UL subframes. Even in normal downlink subframes, mini-slots could be used to transmit NR channels and signals on OFDM symbols not carrying CRS. This, however, may leave almost half the resources of a normal DL subframe unusable for NR, so other techniques may be preferable. For example, symbols carrying CRS could also puncture NR transmissions, similar to URLLC transmissions that pre-empt NR transmissions. The same mechanisms specified for eMBB/URLLC coexistence could be used for LTE/NR coexistence.

Frequency Domain LTE/NR Coexistence Techniques

Frequency domain coexistence between LTE and NR can also be dynamic or semi-static. Semi-static FDM-based coexistence is illustrated in Figure 49. Dynamic frequency domain coexistence is possible when the (e/g) NB schedules both LTE and NR in the same subframe on a Physical Resource Block (PRB) level.

There also exists the possibility of mixing semi-static and dynamic schemes as well as time division multiplexing (TDM)- and frequency division multiplexing (FDM)-based schemes based on the duplex direction. UL resources could be dynamically shared in a TDM fashion, whereas DL resources would be semi-statically configured and frequency division multiplexed between LTE and NR. For example, LTE could operate in paired spectrum, and NR could use LTE UL resources for NR UL transmissions but be configured with a separate DL or dynamic TDD carrier, such as at a higher frequency band. In this scenario, the LTE DL would be semi-statically frequency division multiplexed with NR, but LTE UL resources would need to be dynamically shared between LTE and NR. The semi-statically frequency division multiplexed NR resources could be for DL only or for both DL and UL. For example, it could be beneficial to allow for NR SRS transmissions on the frequency division multiplexed NR-only carrier.

Several issues need to be addressed for the shared LTE UL carrier. For example, if the non-shared NR carrier operates in mmWave spectrum while the shared NR/LTE carrier operates below 6GHz, the UE does not receive NR DL signals that can be used for power control and timing advance of the NR UL transmissions in the shared LTE UL resources. In this case, NR signals may have to be sent in the LTE-only DL resources or, alternatively, the NR-only UE needs to receive and process LTE signals in the LTE-only DL carrier. To avoid NR UEs processing LTE signals or LTE eNBs transmitting NR signals, 3GPP will need to investigate whether the aforementioned problem could be solved by signaling mechanisms. Regardless, further studies are needed to address these issues.

Coordination Requirements for LTE/NR Coexistence

While semi-static techniques identified for coexistence may require minimal coordination, dynamic (for example, per-TTI) sharing can be done by coordinating the LTE and NR transmissions via three different mechanisms:

1. Co-locating the NR and LTE scheduling.

2. Via the X2 interface (or the evolved version of the X2 interface in the new RAN architecture).

3. Over-the-air.
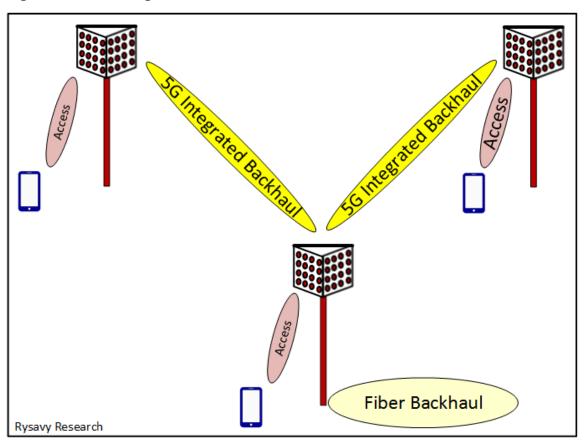
Options 1 and 2 do not impact any RAN1 specification, whereas Option 3 requires RAN1 specifications. Also, over-the-air coordination is desirable because it does not require LTE and NR scheduling and transmission to be handled by a single eNodeB, nor does it require an ultra-low-latency transport between them, thereby providing much more deployment

flexibility. This can even allow NR and LTE to be deployed on different tiers (for example, macro and pico) and share the same channel.

Integrated Access and Backhaul

A study item for Release 15 with the likelihood of standardization in Release 16 is integrated access and backhaul, with which the 5G radio can provide both access and backhaul functions, as shown in Figure 50.

Figure 50: 5G Integrated Access and Backhaul

This 3GPP Study item, which has not yet begun, could address the following capabilities: [126]

- ❑ Wireless backhaul using 5G or LTE.

- ❑ Both indoor and outdoor scenarios.

- ❑ Flexible partitioning between access and backhaul functions.

- ❑ Autonomous configuration of access and wireless self-backhaul functions.

- ❑ Multi-hop wireless self-backhauling to enable flexible extension of range and coverage area.

---

[126] 3GPP, "Study on Integrated Access and Backhaul for NR," RP-170831, March 2017.

❑ Autonomous adaptation on wireless self-backhaul to minimize service disruptions.

❑ Topological redundant connectivity on the wireless self-backhaul to enhance reliability and capacity and to reduce latency.

Performance

5G, with the ability to use wider radio channels than LTE, will be able to deliver much higher peak and average speeds, with initial estimates listed above in the section, "Data Throughput Comparison." In the absence of deployed networks to measure, companies have performed simulations, concentrating initially on one of the first uses cases of fixed wireless access.

Figure 51 shows downlink performance for a network using different base station ISDs, and with and without foliage. The Nokia simulation used base stations with 512 antenna elements, outdoor-mounted user equipment, 28 GHz, and an 800-MHz radio channel, mostly allocated to the downlink.
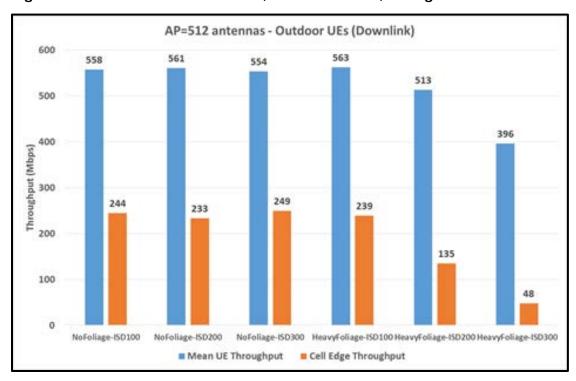
Figure 51: Downlink Performance, Different ISDs, Foliage vs. None[127]

The conclusion of this simulation is that a minimum performance of 100 Mbps at the cell-edge, a 5G objective, is possible at ISDs up to 200 meters, with and without foliage.

The following three figures are from another simulation study by Ericsson, this one also for fixed-wireless access, with the following key assumptions: 350-meter ISD, 96-antenna base stations, 200 MHz radio channels, 57% allocated to downlink, 1000 homes per sq.

---

[127] Nokia contribution. For a full discussion, refer to the associated paper by Frederick W. Vook, Eugene Visotsky, Timothy A. Thomas, and Amitava Ghosh, Nokia Bell Labs, *Performance Characteristics of 5G mmWave Wireless-to-the-Home*, available at http://ieeexplore.ieee.org/document/7869558/.

km., 25% of homes using 4K UHD video service at 15 Mbps, building heights of 4 to 10 meters, and trees from 5 to 15 meters.

Figure 52 shows the throughputs available across the coverage area, with many locations able to receive close to 1 Gbps.
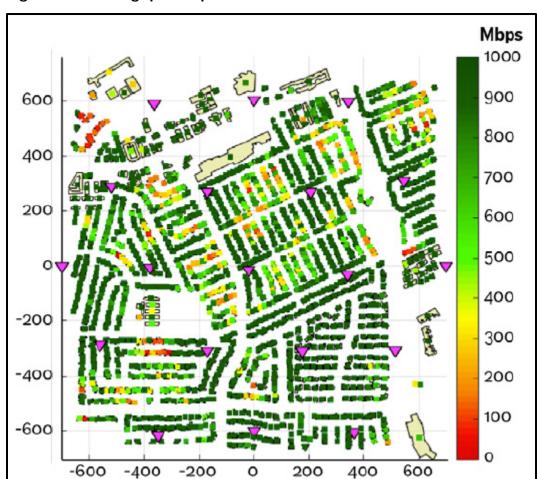
Figure 52: Throughput Map of Suburban Area at Low Load[128]

Figure 53 shows the proportion of users that can obtain 15 Mbps and 100 Mbps service relative to monthly traffic volume. Note that the system supports thousands of GBs of service per subscriber per month.

---

[128] Ericsson contribution, Ericsson Technical Review, *5G and Fixed Wireless Access*, 2016, available at https://www.ericsson.com/assets/local/publications/ericsson-technology-review/docs/2016/etr-5g-and-fixed-wireless-access.pdf.

Figure 53: Proportion of Satisfied Users Relative to Monthly Usage[129]

Figure 54 shows that an ISD of 350 can be used with a combination of indoor, wall-mounted, and rooftop antennas. A large percentage of users, 78%, can use indoor antennas, facilitating deployment.
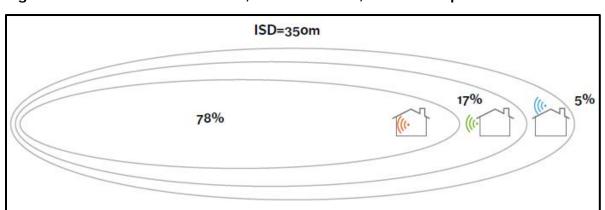
Figure 54: Breakdown of Indoor, Wall-Mounted, and Rooftop Antennas[130]

---

[129] Ibid.

[130] Ibid.

The simulation study shows that 5G fixed wireless access deployments using a larger ISD of 350 meters, translating to 11 base stations per square kilometer, can provide competitive broadband service.

In this environment, handsets with 5G mmWave capability will also be able to access the networks, but the antennas they use may not be as effective as the fixed-wireless equipment, so handsets may need to fall back to 4G, depending on their precise locations. For this reason, the dual-connectivity being planned for 5G will play an important role.

Figure 55 shows another simulation study, this one from Intel, using the following assumptions: 28 GHz operation, 2:1 DL:UL ratio, 25% control overhead, 10 bps/Hz maximum downlink spectral efficiency, CPEs placed either north or south side of house and one with best SNR chosen, and indoor CPE equipment with 30dB outdoor-to-indoor penetration loss. Scenario 1 is 60 access points per sq. km. Scenario 2 is 120 access points per sq. km. (Base grid of 40 houses in a 250x200m area with four rows of 10 houses per row, APs placed along streets and alleys, single-family homes, 4 sectors per AP, and 4.5-meter pole height.)

Figure 55: 5G Fixed Wireless Simulation with Different Loading and Densities[131]

DL	Scenario 1 (3 APs per 40 homes)				Scenario 2 (6 APs per 40 homes)			
	10% loading		50% loading		10% loading		50% loading	
	Cell-edge (Mbps)	Average (Mbps)	Cell-edge (Mbps)	Average (Mbps)	Cell-edge (Mbps)	Average (Mbps)	Cell-edge (Mbps)	Average (Mbps)
100 MHz, 64x8	5	200	4	160	50	290	45	230
100 MHz, 128x16	22	280	20	210	150	370	145	275
400 MHz, 128x16	88	1,120	80	840	600	1,480	580	1,100

UL	Scenario 1 (3 APs per 40 homes)				Scenario 2 (6 APs per 40 homes)			
	10% loading		50% loading		10% loading		50% loading	
	Cell-edge (Mbps)	Average (Mbps)	Cell-edge (Mbps)	Average (Mbps)	Cell-edge (Mbps)	Average (Mbps)	Cell-edge (Mbps)	Average (Mbps)
100MHz, 16x64	2	85	1	75	35	120	30	100
100MHz, 16x128	4	95	2	80	50	140	40	110
100MHz, 32x64	7	100	4	85	70	150	60	115

Using 400 MHz and six access points per 40 homes, and 50% loading, the average throughput was more than 1 Gbps.

LTE and LTE-Advanced

Although HSPA and HSPA+ offer a highly efficient broadband-wireless service that will enjoy success for the remainder of this decade and well into the next, 3GPP completed the specification for Long Term Evolution as part of Release 8. LTE offers even higher peak throughputs in wider spectrum bandwidth. Work on LTE began in 2004 with an official work item started in 2006 and a completed specification early 2009. Initial deployments began in 2010.

LTE uses OFDMA on the downlink, which is well suited to achieve high peak data rates in high-spectrum bandwidth. WCDMA radio technology is basically as efficient as OFDM for delivering peak data rates of about 10 Mbps in 5 MHz of bandwidth. Achieving peak rates

---

[131] Intel contribution.

in the 100 Mbps range with wider radio channels, however, would result in highly complex terminals, and it is not practical with current technology, whereas OFDM provides a practical implementation advantage. Scheduling approaches in the frequency domain can also minimize interference, thereby boosting spectral efficiency. The OFDMA approach is also flexible in channelization: LTE operates in various radio channel sizes ranging from 1.4 to 20 MHz.

On the uplink, however, a pure OFDMA approach results in high Peak to Average Ratio (PAR) of the signal, which compromises power efficiency and, ultimately, battery life. Hence, LTE uses an approach called "SC-FDMA," which is somewhat similar to OFDMA, but has a 2 to 6 dB PAR advantage over the OFDMA method used by other technologies such as WiMAX.

LTE capabilities include:

❑ Downlink peak data rates up to 300 Mbps with 20+20 MHz bandwidth in initial versions, increasing to over 1 Gbps in subsequent versions through carrier aggregation, higher-order modulation, and 4X4 MIMO.

❑ Uplink peak data rates up to 71 Mbps with 20+20 MHz bandwidth in initial versions, increasing to over 1 Gbps in subsequent versions.

❑ Operation in both TDD and FDD modes.

❑ Scalable bandwidth up to 20+20 MHz covering 1.4, 3, 5, 10, 15, and 20 MHz radio carriers.

❑ Increased spectral efficiency over HSPA by a factor of two to four.

❑ Reduced latency, to 15 msec round-trip times between user equipment and the base station, and to less than 100 msec transition times from inactive to active.

❑ Self-organizing capabilities under operator control and preferences that will automate network planning and will result in lower operator costs.

LTE-Advanced Terminology

LTE-Advanced, as specified in Release 10, is a term used for the version of LTE that addresses IMT-Advanced requirements. The ITU ratified LTE-Advanced as IMT-Advanced in November 2010. LTE-Advanced is both backward- and forward-compatible with LTE, meaning LTE devices operate in newer LTE-Advanced networks, and LTE-Advanced devices operate in older, pre-Release 10 LTE networks.

The following lists at a high level the most important features of LTE-Advanced, as well as other features planned for subsequent releases, including Release 11:

❑ Carrier aggregation.

❑ Higher-order downlink MIMO (up to 8X8 in Release 10).

❑ Uplink MIMO (two transmit antennas in the device).

❑ Coordinated multipoint transmission (CoMP) in Release 11.

❑ Heterogeneous network (HetNet) support including Enhanced Inter-cell Interference Coordination (eICIC).

❑ Relays.

3GPP, from Release 13, has referred to LTE as LTE-Advanced Pro, which includes features such as LAA, LWA, low latency, and massive MIMO.
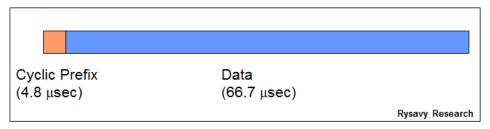
OFDMA and Scheduling

LTE implements OFDM in the downlink. The basic principle of OFDM is to split a high-rate data stream into a number of parallel, low-rate data streams, each a narrowband signal carried by a subcarrier. The different narrowband streams are generated in the frequency domain, and then combined to form the broadband stream using a mathematical algorithm called an "Inverse Fast Fourier Transform" (IFFT) that is implemented in digital signal processors. In LTE, the subcarriers have 15 kHz spacing from each other. LTE maintains this spacing regardless of the overall channel bandwidth, which simplifies radio design, especially in supporting radio channels of different widths. The number of subcarriers ranges from 72 in a 1.4 MHz radio channel to 1,200 in a 20 MHz radio channel.

The composite signal obtained after the IFFT is extended by repeating the initial part of the signal (called the Cyclic Prefix [CP]). This extended signal represents an OFDM symbol. The CP is basically a guard time during which reflected signals will reach the receiver. It results in an almost complete elimination of multipath-induced Intersymbol Interference (ISI), which otherwise makes extremely high data rate transmissions problematic. The system is called orthogonal because the subcarriers are generated in the frequency domain (making them inherently orthogonal), and the IFFT conserves that characteristic.

OFDM systems may lose their orthogonal nature as a result of the Doppler shift induced by the speed of the transmitter or the receiver. 3GPP specifically selected the subcarrier spacing of 15 kHz to avoid any performance degradation in high-speed conditions. WiMAX systems that use a lower subcarrier spacing (~11 kHz) are more impacted in high-speed conditions than LTE.

Figure 56: OFDM Symbol with Cyclic Prefix

The multiple access aspect of OFDMA comes from being able to assign different users different subcarriers over time. A minimum resource block that the system can assign to a user transmission consists of 12 subcarriers over 14 symbols in 1.0 msec. Figure 57 shows how the system can assign these resource blocks to different users over both time and frequency.
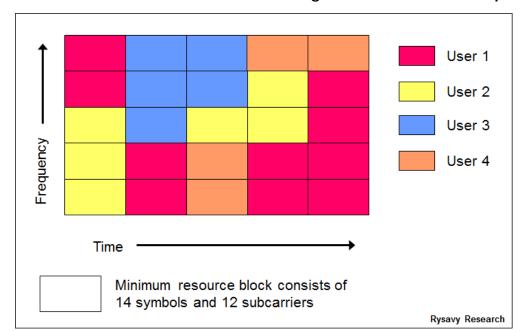
Figure 57: LTE OFDMA Downlink Resource Assignment in Time and Frequency

By controlling which subcarriers are assigned in which sectors, LTE can easily control frequency reuse. Using all the subcarriers in each sector, the system would operate at a frequency reuse of 1; but by using a different one third of the subcarriers in each sector, the system can achieve a looser frequency reuse of 1/3. The looser frequency reduces overall spectral efficiency but delivers high peak rates to users.

Beyond controlling frequency reuse, frequency domain scheduling, as shown in Figure 58 can use those resource blocks that are not faded, not possible in CDMA-based systems. Since different frequencies may fade differently for different users, the system can allocate those frequencies for each user that result in the greatest throughput. This results in up to a 40% gain in average cell throughput for low user speed (3 km/hour), assuming a large number of users and no MIMO. The benefit decreases at higher user speeds.
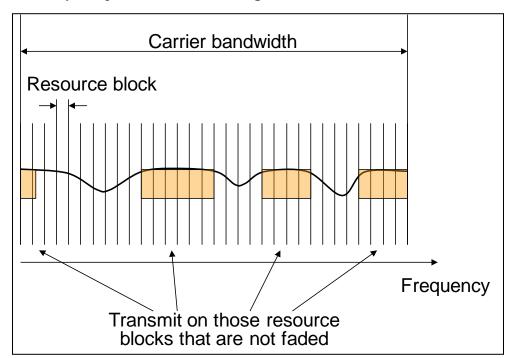
Figure 58: Frequency Domain Scheduling in LTE[132]

LTE Smart Antennas

Wireless networks can achieve significant gains by employing multiple antennas, either at the base station, the mobile device, or both. LTE uses multiple antennas in three fundamentally different ways:

- ❑ **Diversity.** So long as the antennas are spaced or polarized appropriately, the antennas provide protection against fading.

- ❑ **Beamforming.** Multiple antennas can shape a beam to increase the gain for a specific receiver. Beamforming can also suppress specific interfering signals. Beamforming is particularly helpful for improving cell-edge performance.

- ❑ **Spatial Multiplexing.** Often referred to as MIMO antenna processing, spatial multiplexing creates multiple transmission paths through the environment, effectively sending data in parallel through these paths, thus increasing both throughput and spectral efficiency.

Table 18 shows the varous antenna transmission modes.

---

[132] 5G Americas member contribution.

Table 18: LTE Transmission Modes[133]

Transmission Mode	Description
1	Single antenna transmission.
2	Transmit Diversity.
3	Transmit diversity for one layer, open-loop codebook-based precoding if more than one layer.
4	Closed-loop codebook-based precoding.
5	Multi-user MIMO version of transmission mode 4.
6	Special case of closed-loop codebook-based precoding limited to single layer transmission.
7	Beamforming. (Non-codebook-based precoding supporting one layer.)
8	Dual-layer beamforming. (Release 9. Non-codebook-based precoding supporting up to two layers.)
9	8-layer transmission. (Release 10. Non-codebook-based precoding supporting up to eight layers.)
10	8-layer transmission with support for CoMP. (Release 11.)

Being able to exploit different antenna modes based on local conditions produces huge efficiency and performance gains, and is the reason that 3GPP is developing even more advanced antenna modes in subsequent LTE releases.

Precoding refers to a mathematical matrix operation performed on radio symbols to determine how they are combined and mapped onto antenna ports. The precoder matrix can operate in either open-loop or closed-loop modes. For each transmission rank for a given number of transmission ports (antennas), there is a limited set of precoder matrices defined, called the codebook. This helps limit the amount of signaling needed on uplink and downlink.

Fundamental variables distinguish the different antenna modes:

❑ **Single base station antenna versus multiple antennas.** Single antennas provide for Single Input Single Output (SISO), SIMO, and planar-array beamforming. (Multiple Output means the UE has multiple antennas.) Multiple antennas at the base station provide for different MIMO modes such as 2X2, 4X2, and 4X4.

---

[133] Erik Dahlman, Stefan Parkvall, Johan Skold, *4G - LTE/LTE Advanced for Mobile Broadband*, Academic Press, 2011.

- **Single-user MIMO versus multi-user MIMO.** Release 8 only provides for single-user MIMO on the downlink. Release 10 includes multi-user MIMO.

- **Open-Loop versus Closed-Loop.** High vehicular speeds require open-loop operation whereas slow speeds enabled closed-loop operation in which feedback from the UE modifies the transmission. In closed-loop operation, the precoder matrix is based on this feedback.

- **Rank.** In a MIMO system, the channel rank is formally defined as the rank of the channel matrix and is a measure of the degree of scattering that the channel exhibits. For example, in a 2x2 MIMO system, a rank of one indicates a low-scattering environment, while a rank of two indicates a high-scattering environment. The rank two channel is highly uncorrelated, and is thus able to support the spatial multiplexing of two data streams, while a rank one channel is highly correlated, and thus can only support single stream transmission (the resulting multi-stream interference in a rank one channel as seen at the receiver would lead to degraded performance). Higher Signal to Interference plus Noise Ratios (SINR) are typically required to support spatial multiplexing, while lower SINRs are typically sufficient for single stream transmission. In a 4x4 MIMO system channel rank values of three and four are possible in addition to values of one and two. The number of data streams, however, or more specifically codewords in LTE is limited to a value of two. Thus, LTE has defined the concept of layers, in which the DL transmitter includes a codeword-to-layer mapping, and in which the number of layers is equal to the channel rank. An antenna mapping or precoding operation follows, which maps the layers to the antenna ports. A 4x2 MIMO system is also possible with LTE Release 8, but here the channel rank is limited to the number of UE antennas, which is equal to two.

The network can dynamically choose between different modes based on instantaneous radio conditions between the base station and the UE. Figure 59 shows the decision tree. The antenna configuration (AC) values refer to the transmission modes. Not every network will support every mode. Operators will choose which modes are the most effective and economical. AC2, 3, 4, and 6 are typical modes that will be implemented.
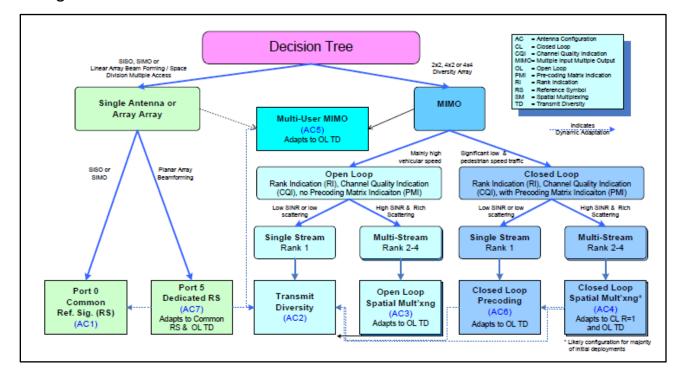
Figure 59: Decision Tree for Different Antenna Schemes[134]

The simplest mode is AC2, referred to as Transmit Diversity (TD) or sometimes Space Frequency Block Code (SFBC) or even Open Loop Transmit Diversity. TD can operate under all conditions, meaning it works under low SINR, high mobility, and low channel rank (rank = 1). This rank means that the channel is not sufficiently scattered or de-correlated to support two spatial streams. Thus, in TD, only one spatial stream or what is sometimes referred as a single codeword (SCW) is transmitted. If the channel rank increases to a value of two, indicating a more scattered channel, and the SINR is a bit higher, then the system can adapt to AC3 or Open-Loop Spatial Multiplexing (OL-SM), also referred to as large-delay Cyclic Delay Diversity (CDD). This mode supports two spatial streams or two codewords. This mode, also called multiple codeword (MCW) operation, increases throughput over SCW transmission.

If the rank of the channel is one, but the device is not moving very fast or is stationary, then the system can adapt to AC6, called closed-loop (CL) precoding (or CL-rank 1 or CL-R1). In this mode, the network receives from the device with Precoding Matrix Indication (PMI) bits that inform the base station what precoding matrix to use in the transmitter to optimize link performance. This feedback is only relevant for low-mobility or stationary conditions since in high mobility conditions the feedback will most likely be outdated by the time the base station can use it.

Another mode is AC4 or Closed Loop Spatial Multiplexing (CL-SM), which is enabled for low-mobility, high SINR, and channel rank of two. This mode theoretically provides the best user throughput. The figure above shows how these modes can adapt downwards to either OL TD, or if in CL-SM mode, down to either OL TD or CL R1.

[134] 4G Americas *MIMO and Smart Antennas for 3G and 4G Wireless Systems – Practical Aspects and Deployment Considerations*, May 2010.

For a 4x4 MIMO configuration, the channel rank can take on values of three and four in addition to one or two. Initial deployment at the base station, however, will likely be two TX antennas and most devices will only have 2 RX antennas, and thus the rank is limited to 2.

AC5 is MU-MIMO, which is not defined for the downlink in Release 8.

AC1 and AC7 are single antenna port modes in which AC1 uses a common Reference Signal (RS), while AC7 uses a dedicated RS or what is also called a user specific RS. AC1 implies a single TX antenna at the base station. AC7 implies an antenna array with antennal elements closely spaced so that a physical or spatial beam can be formed toward an intended user.

LTE operates in a variety of MIMO configurations. On the downlink, these include 2X2, 4X2 (four antennas at the base station), and 4X4. Initial deployment will likely be 2x2 whereas 4X4 will be most likely used initially in femtocells. On the uplink, there are two possible approaches: single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO). SU-MIMO is more complex to implement as it requires two parallel radio transmit chains in the mobile device, whereas MU-MIMO does not require any additional implementation at the device but relies on simultaneous transmission on the same tones from multiple mobile devices.

The first LTE Release thus incorporates MU-MIMO with SU-MIMO deferred for subsequent LTE releases. An alternate form of MIMO, originally called network MIMO, and now called CoMP, relies on MIMO implemented (on either the downlink or uplink or both) using antennas across multiple base stations, as opposed to multiple antennas at the same base station. This paper explains CoMP in the section on LTE Advanced below.

Peak data rates are approximately proportional to the number of send and receive antennas. 4X4 MIMO is thus theoretically capable of twice the data rate of a 2X2 MIMO system. The spatial multiplexing MIMO modes that support the highest throughput rates will be available in early deployments.

For a more detailed discussion of 3GPP antenna technologies, refer to the 5G Americas white paper "MIMO and Smart Antennas for 3G and 4G Wireless Systems – Practical Aspects and Deployment Considerations," May 2010.

For advancements in LTE Smart Antennas, see the next section.

LTE-Advanced Antenna Technologies

Release 10 added significant enhancements to antenna capabilities, including four-layer transmission resulting in peak spectral efficiency exceeding 15 bps/Hz. Uplink techniques fall into two categories: those relying on channel reciprocity and those that do not. With channel reciprocity, the eNB determines the channel state by processing a Sounding Reference Signal from the UE. It then forms transmission beams accordingly. The assumption is that the channel received by the eNB is the same as the UE. Techniques that use channel reciprocity are beamforming, SU-MIMO, and MU-MIMO. Channel reciprocity works especially well with TDD since both forward and reverse links use the same frequency.

Non-reciprocity approaches apply when the transmitter has no knowledge of the channel state. Techniques in this instance include open-loop MIMO, closed-loop MIMO, and MU-MIMO. These techniques are more applicable for higher speed mobile communications.

For the downlink, the technology can transmit in as many as eight layers using an 8X8 configuration for a peak spectral efficiency of 30 bps/Hz. This exceeds the IMT-Advanced requirements, conceivably supporting a peak rate of 1 Gbps in just 40+40 MHz, and even higher rates in wider bandwidths. This would require additional reference signals for channel estimation and for measurements, including channel quality, to enable adaptive, multi-antenna transmission.

Release 10 supports a maximum of two codewords, the same as previous LTE releases. The release specifies a new transmission mode (TM-9) that supports SU-MIMO up to Rank 8 (up to eight layers), as well as the ability to dynamically switch between SU-MIMO and MU-MIMO.

Figure 60 shows the different forms of single-user MIMO in Releases 8, 9, and 10. Release 8 supports only a single layer, whereas two-layer beamforming is possible in Release 9, and eight layers are possible in Release 10 with eight antennas at the base station.

Figure 60: Single-User MIMO[135]
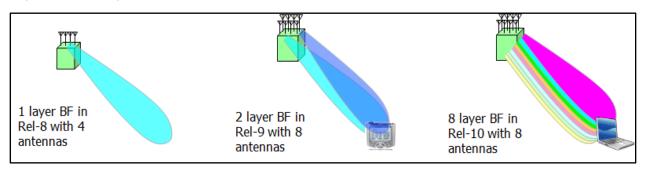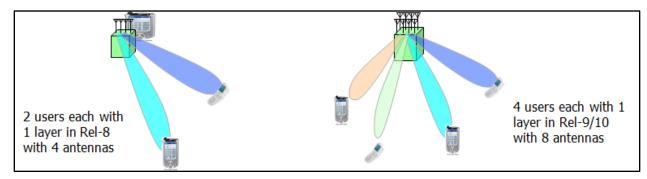
Figure 61 shows multi-user MIMO options across different releases. Release 8 supports two simultaneous users, each with one layer using four antennas, while Releases 9 and 10 support four simultaneous users, each with one layer.

Figure 61: Multi-User MIMO[136]

For four-antenna configurations at the base station, Release 12 improves throughput by adding a feedback mode, called mode 3-2, in which sub-band precoders and sub-band

---

[135] 5G Americas member contribution.

[136] 5G Americas member contribution.

channel quality indicators (CQIs) are included in the UE's feedback to the eNodeB. Release 12 also adds a new codebook that further improves throughput.

As depicted in Figure 62 and Figure 63, compared with the Release 8 codebook, the new Release 12 codebook provides a 10% gain for both median and cell-edge throughputs. Compared with feedback mode 3-1, feedback mode 3-2 provides an 18% to 20% gain in median and cell-edge throughput. Jointly, the two methods provide a 28% to 30% gain.

Figure 62: Median Throughput of Feedback Mode 3-2 and New Codebook.[137]
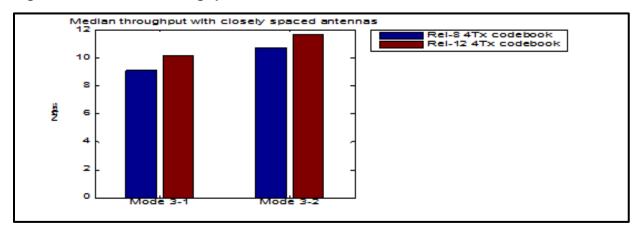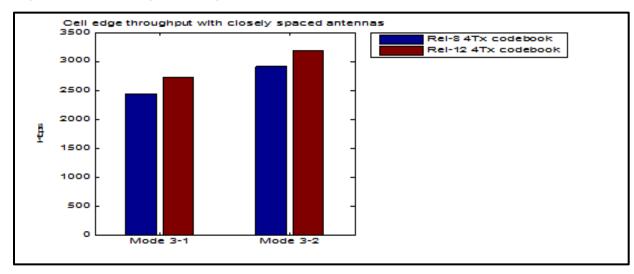
Figure 63: Cell-Edge Throughput of Feedback Mode 3-2 and New Codebook[138]

Release 12 also defines how Active Antenna Systems can use multiple transceivers on an antenna array to dynamically adjust a radiation pattern.

---

[137] 5G Americas member contribution. Assumptions include: cellular layout of 19 sites hexagonal grid with three sectors per site and 500-meter inter-site distance; simulation case ITU uMa for macro; carrier frequency 2 GHz, deployment scenario A homogenous macro; SU-MIMO with maximum two layers per UE; proportional fair scheduler; and bursty traffic model.

[138] 5G Americas member contribution. Same assumptions as previous figure.

Release 13 is likely to define full-dimension MIMO, which adds a large number of antenna elements, potentially as many as 64 elements.

A practical consideration with antennas is that many towers today already support multiple operators, with tower companies having to manage interference placement, spectrum allocations, and wind and snow load. At higher frequencies, a single radome (antenna enclosure) can support 4X2 MIMO, but higher-order MIMO may prove impractical for many deployments.

5G systems operating at much higher frequencies will have an advantage since the antenna arrays will be much smaller due to the much smaller wavelengths.

Initial massive MIMO techniques applied to LTE, such as full-dimension MIMO using 8, 16, and 64 transmit antennas, can provide dramatic performance gains, particularly in dense deployments, as shown in Figure 64.
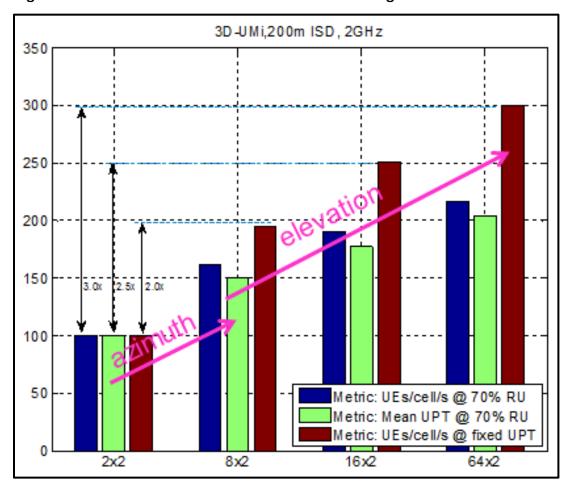
Figure 64: Performance Gains with FD-MIMO Using 200 Meter ISD[139]

This figure compares 8X2, 16X2, and 64X2 MIMO performance relative to 2X2 MIMO (normalized to value 100). The blue bars (case 1) show the supported number of users per sector (referred to as "cell" in the figure) at a fixed resource utilization (RU) of 70%; the

---

[139] 5G Americas member contribution.

green bars (case 2) show mean user throughput (UPT) at a fixed RU of 70%; and the red bars (case 3) show system capacity in terms of supported number of users for a given user throughput. Resulting gains are:

- ❑ Case 2 (green bars): 1.5X with 8X2, 1.75X with 16X2, and 2X with 64X2 MIMO.

- ❑ Case 3 (red bars): 2X with 8X2, 2.5X with 16X2, and 3X with 64X2 MIMO.

The primary gains are from azimuth (horizontal dimension) in going from 2X2 to 8X2, and from elevation in going to 16X2 and 64X2. FD-MIMO gains are lower with larger ISD values, such as 500 meters.

3GPP has also studied FD-MIMO and conducted a field trial showing impressive throughput gains, particularly in a high-rise scenario.[140]

Carrier Aggregation

Carrier aggregation, first available in Release 10, plays an important role in providing operators maximum flexibility for using all of their available spectrum. By combining spectrum blocks, LTE can deliver much higher throughputs than otherwise possible. Asymmetric aggregation (for example, different amounts of spectrum used on the downlink versus the uplink) provides further flexibility and addresses the greater demand on downlink traffic.

Specific types of aggregation include:

- ❑ Intra-band on adjacent channels.

- ❑ Intra-band on non-adjacent channels.

- ❑ Inter-band (700 MHz, 1.9 GHz).

- ❑ Inter-technology (for example, LTE on one channel, HSPA+ on another). This approach is not currently specified nor being developed. While theoretically promising, a considerable number of technical issues would have to be addressed.[141] See Figure 65.

---

[140] 3GPP, *3D-MIMO Prototyping and Initial Field Trial Results*, TSG RAN WG1 Meeting #80, Agenda Item: 7.2.4.4, Document R1-150451.

[141] For further details, see 4G Americas, *HSPA+ LTE Carrier Aggregation*, June 2012.

Figure 65: Inter-Technology Carrier Aggregation[142]

Figure 66 depicts the carrier-aggregation capabilities of different 3GPP releases.

---

[142] 5G Americas member contribution.

Figure 66: Carrier Aggregation Capabilities across 3GPP Releases[143]

One anticipated benefit of inter-band aggregation stems from using the lower-frequency band for users who are at the cell edge, to boost their throughput rates. Though this approach improves average aggregate throughput of the cell by only a small amount (say, 10%), it results in a more uniform user experience across the cell coverage area.

Figure 67 shows an example of intra-band carrier aggregation using adjacent channels with up to 100+100 MHz of bandwidth supported. Radio-access network specifications, however, limit the number of carriers to two in Release 10 and Release 11.
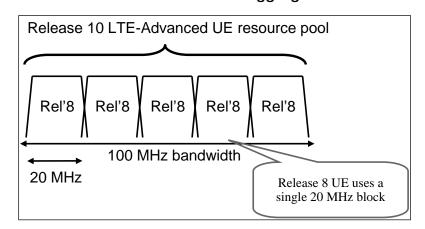
Figure 67: Release 10 LTE-Advanced Carrier Aggregation[144]

---

[143] 4G Americas, *Mobile Broadband Evolution: Rel-12 & Rel-13 and Beyond*, 2015.

[144] Harri Holma and Antti Toskala, *LTE for UMTS, OFDMA and SC-FDMA Based Radio Access*, Wiley, 2009.

Figure 68 shows the carrier aggregation operating at different protocol layers.

Figure 68: Carrier Aggregation at Different Protocol Layers[145]

For a list of band combinations, refer to the 5G Americas white paper, *Wireless Technology Evolution Towards 5G: 3GPP Release 13 to Release 15 and Beyond,* February 2017, at section 3.4.3.

Figure 69 shows the result of one simulation study that compares download throughput rates between the blue line, which shows five user devices in 700 MHz and five user devices in AWS not using CA, and the pink line, which shows ten user devices that have access to both bands. Assuming a lightly loaded network with CA, 50% or more users (the median) experience 91% greater throughput, and 95% or more users experience 50% greater throughput. These trunking gains are less pronounced in heavily loaded networks.

---

[145] Stefan Parkvall and David Astely, Ericsson Research, "The Evolution of LTE towards IMT-Advanced," Journal of Communications, Vol. 4, No. 3, April 2009. Available at http://www.academypublisher.com/jcm/vol04/no03/jcm0403146154.pdf.
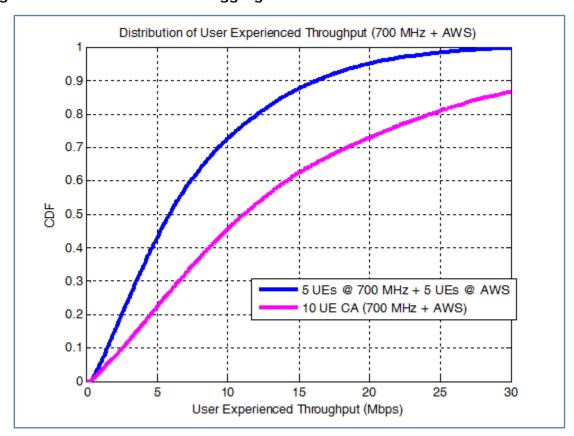
Figure 69: Gains from Carrier Aggregation[146]

Distribution of User Experienced Throughput (700 MHz + AWS)

Work in Release 12 is investigating aggregation of joint TDD and FDD carriers.

Coordinated Multi Point (CoMP)

Coordinated Multi Point (CoMP) is a communications technique that can improve coverage, cell-edge throughput, and/or system spectrum efficiency by reducing interference. This technique was thoroughly studied during the development of LTE-Advanced Release 10 and was standardized in Release 11.

CoMP coordinates transmissions at different cell sites, thereby achieving higher system capacity and improving cell-edge data rates.

The main principle of CoMP is that a UE at a cell edge location can receive signals from multiple transmission points, and/or its transmitted signal can be received by multiple reception points. Consequently, if these multiple transmission points coordinate their transmissions, the DL throughput performance and coverage can improve.
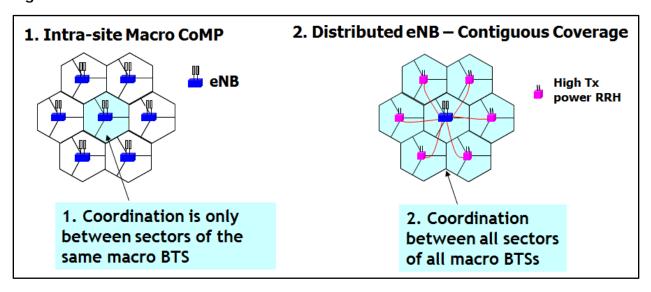
For the UL, signals from the UE received at multiple reception points can significantly improve the link performance. Techniques can range from simple interference avoidance methods, such as Coordinated Beam Switching (CBS) and Coordinated Beam Forming

---

[146] 5G Americas member contribution. Assumptions: lightly-loaded network, 2.0 site-to-site distance, file size is 750 Kbytes, traffic model bursty with mean inter-arrival time of five seconds.

(CBF), to complex joint processing techniques that include Joint Transmission (JT), Joint Reception (JR), and Dynamic Point Selection (DPS).
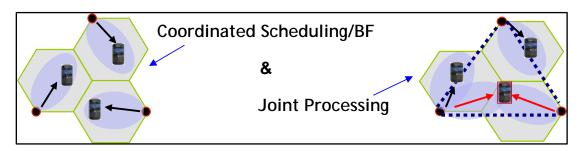
CoMP architectures include inter-site CoMP, intra-site CoMP, as well as CoMP with distributed eNBs (i.e., an eNB with distributed remote radio heads). Figure 70 shows two possible levels of coordination.

Figure 70: Different Coordination Levels for CoMP[147]

In one CoMP approach, called coordinated scheduling and shown in Figure 71, a single site transmits to the user, but with scheduling, including any associated beamforming, coordinated between the cells to reduce interference between the different cells and to increase the served user's signal strength. In Joint Transmission, another CoMP approach also shown in Figure 71, multiple sites transmit simultaneously to a single-user. This approach can achieve higher performance than coordinated scheduling, but it has more stringent backhaul communications requirements. One simpler form of CoMP that will be available in Release 10, and then further developed in Release 11, is ICIC. Release 11 of LTE defines a common feedback and signaling framework for enhanced CoMP operation.

Figure 71: Coordinated Scheduling/BF and Joint Processing CoMP Approaches[148]

---

[147] 5G Americas member contribution.

[148] 5G Americas member contribution.

Release 11 also implements CoMP on the uplink, by which multiple base stations receive uplink transmissions and jointly process the signal, resulting in significant interference cancellation and improvements to spectral efficiency.

The performance gains expected from CoMP are under discussion in the industry. According to 3GPP document TR 36.819, for the case of resource utilization below 35%, CoMP may provide a 5.8% performance gain on the downlink for the mean user and a 17% gain for cell-edge users relative to HetNets without eICIC. For resource utilization of more than 35%, CoMP may provide a 17% mean gain and a 40% cell-edge gain.[149] CoMP can also be used in combination with eICIC for additional gains.

In the same 3GPP TR 36.819 document, 3GPP estimates the downlink CoMP gain in spectral efficiency, defined as average sector throughput for full buffer traffic using JT and 4x2 MU-MIMO as defined in R11, compared with 4x2 MU-MIMO based on R10, to be about 3% for intra-eNodeB CoMP. That gain drops to about 9% for inter-eNodeB CoMP in the case of no delay in the backhaul used to exchange information between eNodeBs. The corresponding gains in cell-edge user throughput are 20% and 31%, respectively.

When increasing the backhaul latency to a more realistic value of 10 msec for inter-eNodeB, spectral efficiency decreases to zero, and the cell edge gain decreases to 10%.

The gains for DL CoMP based on Coordinated Scheduling/Coordinated Beamforming (CS/CB) and intra-eNodeB are less than that provided by JT, with spectral efficiency at 1% and cell edge gains at 4%.

All of the above gains are for FDD networks with cross-polarized antennas at the eNodeBs. For TDD networks, the gains are higher by virtue of being able to invoke channel reciprocity and thus infer the DL channel directly from the UL channel. For example, for intra-eNodeB CoMP with JT 4x2 MU-MIMO, the respective gains in spectral efficiency and cell-edge throughput are 14% and 29%, respectively.

The gains for UL CoMP based on Joint Reception (JR) are greater than the DL gains. For intra-eNodeB CoMP, the average and cell-edge throughputs are increased to 22% and 40%, assuming two receive antenna paths with SU-MIMO. These respective gains increase to 31% and 66% for inter-eNodeB CoMP. In addition, UL CoMP does not require standardization and thus facilitates vendor implementation.

Uplink CoMP assists VoLTE because it improves cell-edge performance, making voice handover more reliable when traversing between cells. The benefit is analogous to CDMA soft handover; in both cases, the mobile device communicates with two sites simultaneously.

Cellular V2X Communications

In Release 14, 3GPP is specifying cellular vehicle-to-X (C-V2X) communications with two complementary transmission modes: direct communications between vehicles and network communications.

Direct communications will use bands such as the Intelligent Transportation Systems (ITS) 5.9 GHz band, using the PC5 interface specified for LTE device-to-device communications,

---

[149] 3GPP, *Coordinated Multi-Point Operation for LTE Physical Layer Aspects, TR 36.819 v11.1.0,* Tables 7.3.1.2-3 and 7.3.1.2-4, September 2011.

and will not require a Universal Integrated Circuit Card (UICC) SIM (USIM) card. By operating on different channels in the ITS band, direct cellular V2X will be able to co-exist with IEEE 802.11p, another automotive communications protocol.

In the network communications mode, the system will use traditional cellular licensed spectrum.

Use cases include do-not-pass warnings, blind-curve hazard warnings, road-works warnings, blind-intersection assistance, bicyclist and pedestrian alerts, and left-turn assistance.

C-V2X is being designed to be compatible with other automotive standards, such as those from ETSI and the Society of Automotive Engineers.

User-Plane Congestion Management (UPCON)

With User-Plane Congestion Management, specified in Release 13, operators have additional tools to mitigate network congestion in specific coverage areas. Mechanisms include traffic prioritization by adjusting QoS for specific services; reducing traffic by, for example, compression; and limiting traffic, such as by prohibiting or deferring certain traffic.

3GPP specifications add a new architectural entity, called the "RAN Congestion Awareness Function" (RCAF), that determines whether a cell is congested, determines the UEs supported by that cell, and informs the Policy Control and Charging Rules Function (PCRF), which can subsequently apply different policies to mitigate the congestion.[150]

Network-Assisted Interference Cancellation and Suppression (NAICS)

NAICS, a Release 13 capability, enhances the interference cancellation and suppression capability of UEs by using more information from the network. The fundamental goal of NAICS is to identify and cancel the dominant interferer, not an easy task when the dominant interferer can be on or off and can change in time and frequency. One analysis estimates an average performance gain of 7.4% relative to Release 11 Interference Rejection Combining and 11.7% at the cell edge.[151] 5G Americas members expect even higher performance gains, for example 20%, with implementation-specific scheduling and as NAICS methods are refined.

Multi-User Superposition Transmission (MUST)

MUST, a study item in Release 13 and tentatively planned for Release 14 uses simultaneous transmissions of data for more than one UE within a cell without time, frequency, or spatial layer separation. The concept relies on a UE close to the base station having low propagation loss and a UE far from the base station having high propagation loss. The far UE is not aware of, nor interfered by the near UE transmission. The near UE cancels the far UE interference. The capacity gain grows with the SNR/SINR difference between the close and far UEs.

---

[150] For further details, see 3GPP TR 23.705, "Study on system enhancements for user plane congestion management (Release 13)."

[151] Harri Holma, Antti Toskala, Jussi Reunanen, *LTE Small Cell Optimization: 3GPP Evolution to Release 13*, Jan 2016, Wiley, ISBN: 978-1-118-91257-7.

IPv4/IPv6

Release 8 defines support for IPv6 for both LTE and UMTS networks. An Evolved Packet System bearer can carry both IPv4 and IPv6 traffic, enabling a UE to communicate both IPv4 and IPv6 packets (assuming it has a dual stack) while connected through a single EPS bearer. It is up to the operator, however, whether to assign IPv4, IPv6, or both types of addresses to UE.

Communicating between IPv6-only devices and IPv4 endpoints will require protocol-conversion or proxies. For further details, refer to the 5G Americas white paper, "IPv6 – Transition Considerations for LTE and Evolved Packet Core," February 2009.

TDD Harmonization

3GPP developed LTE TDD to be fully harmonized with LTE FDD including alignment of frame structures, identical symbol-level numerology, the possibility of using similar Reference Signal patterns, and similar synchronization and control channels. Also, there is only one TDD variant. Furthermore, LTE TDD has been designed to co-exist with TD-SCDMA and TD-CDMA/UTRA (both low-chip rate and high-chip rate versions). LTE TDD achieves compatibility and co-existence with TD-SCDMA by defining frame structures in which the DL and UL time periods can be time aligned to prevent BTS to BTS and UE to UE interference to support operation in adjacent carriers without the need for large guardbands between the technologies. This will simplify deployment of LTE TDD in countries such as China that are deploying TD-SCDMA. Figure 72 demonstrates the synchronization between TC-SCDMA and LTE-TDD in adjacent channels.
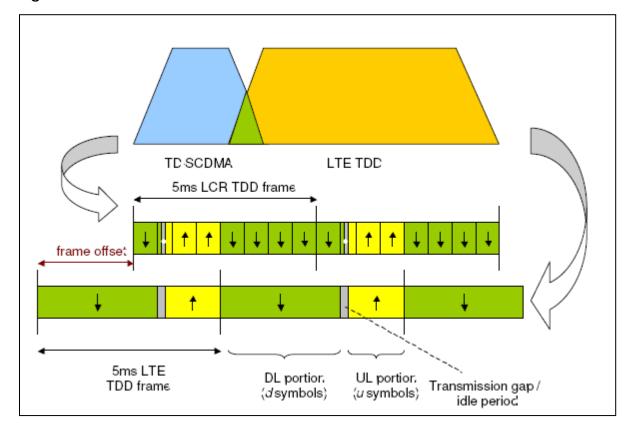
Figure 72: TDD Frame Co-Existence between TD-SCDMA and LTE TDD[152]

For LTE FDD and TDD to co-exist, large guardbands will be needed to prevent interference.

SMS in LTE

Even if an LTE network uses CSFB for voice, LTE devices will be able to send and receive SMS messages while on the LTE network. In this case, the 2G/3G core network will handle SMS messaging, but will tunnel the message to the MME in the EPC via the SGs interface. Once an LTE network uses IMS and VoLTE for packet voice service, SMS will be handled as SMS over IP and will use IMS infrastructure.[153]

User Equipment Categories

LTE specifications define categories of UE, which mainly determine the maximum throughputs of devices but also govern the number of downlink MIMO layers, as shown in Table .

Higher throughput capabilities are possible with 64 QAM and 256 QAM modulation. 3GPP is also defining Category 0 and Category M devices for M2M, as discussed in the section "Internet of Things and Machine-to-Machines."

---

[152] 5G Americas member company contribution.

[153] For further details, see 4G Americas, *Coexistence of GSM, HSPA and LTE*, May 2011, 35.

Table 19: UE Categories[154]

UE Category	Max DL Throughput	Maximum DL MIMO Layers	Maximum UL Throughput
1	10.3 Mbps	1	5.2 Mbps
2	51.0 Mbps	2	25.5 Mbps
3	102.0 Mbps	2	51.0 Mbps
4	150.8 Mbps	2	51.0 Mbps
5	299.6 Mbps	4	75.4 Mbps
6	301.5 Mbps	2 or 4	51.0 Mbps
7	301.5 Mbps	2 or 4	102.0 Mbps
8	2998.6 Mbps	8	1497.8 Mbps
9	452.3 Mbps	2 or 4	51.0 Mbps
10	452.3 Mbps	2 or 4	102.0 Mbps
11	603.0 Mbps	2 or 4	51.0 Mbps
12	603.0 Mbps	2 or 4	102.0 Mbps
13	391.6 Mbps	2 or 4	150.8 Mbps
14	3916.6 Mbps	8	9587.7 Mbps
15	798.8 Mbps	2 or 4	226.1 Mbps
16	1051.4 Mbps	2 or 4	105.5 Mbps
17	2506.6 Mbps	8	2119.4 Mbps
18	1206.0 Mbps	2 or 4 (or 8)	211.0 Mbps
19	1658.3 Mbps	2 or 4 (or 8)	13563.9 Mbps
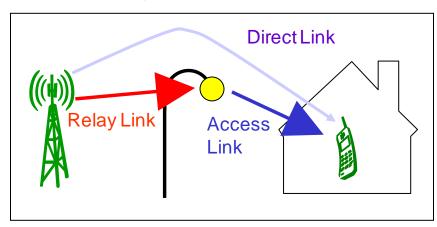
LTE-Advanced Relays

Another capability being planned for LTE-Advanced is relays, as shown in Figure 73. The idea is to relay frames at an intermediate node, resulting in much better in-building penetration, and with better signal quality, user rates will improve. Relay nodes can also improve cell-edge performance by making it easier to add picocells at strategic locations.

---

[154] 3GPP, E*volved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio access capabilities,* Technical Specification 36.306 V14.2.0, March 2017.

Relays provide a means for lowering deployment costs in initial deployments in which usage is relatively low. As usage increases and spectrum needs to be allocated to access only, operators can then employ alternate backhaul schemes.

Figure 73: LTE-Advanced Relay[155]

Proximity Services (Device-to-Device)

Release 12 defined a capability for devices to communicate directly with one another using LTE spectrum, a feature also called "operator-enabled proximity services." With this capability, devices can autonomously discover nearby relevant devices and services in a battery-efficient manner. Devices broadcast their needs and services and can also passively identify services without user intervention. The communication between devices is called "sidelink communications" and uses an interface called "PC5." Release 12, emphasizing public-safety applications, supports only one-to-many sidelink communications, whereas Release 13 supports one-to-one sidelink communications between two group member UEs and between a remote UE and a relay UE.

Initial emphasis of this capability, in both Release 12 and Release 13, is on public safety. Examples of potential consumer or commercial applications include discovering friends and family (social matching), push advertising for relevant notifications, tourist bulletins, venue services, crime alerts, home automation, vehicle-to-vehicle communication, and detecting children leaving the vicinity of their homes. The service is designed to work during infrastructure failures, even in emergencies and natural disasters. As a new means of communicating, proximity services could result in innovative types of applications.

The LTE network performs configuration and authentication; however, communication can be either via the network or directly between devices. To minimize battery consumption, devices synchronously wake up for brief intervals to discover services. The impact on LTE network capacity is minimal.

As with other location-based services, operators and application developers will need to address privacy concerns.

---

[155] 5G Americas member contribution.

LTE Throughput

The section "4G LTE Advances" above in the main section of the paper and "Data Throughput Comparison" in the appendix provide an overview of LTE throughputs. This section provides additional details.

Table shows initial (Release 8) LTE peak data rates based on different downlink and uplink designs.

Table20: LTE Peak Throughput Rates

LTE Configuration	Downlink (Mbps) Peak Data Rate	Uplink (Mbps) Peak Data Rate
Using 2X2 MIMO in the Downlink and 16 QAM in the Uplink, 10+10 MHz	70.0	22.0
Using 4X4 MIMO in the Downlink and 64 QAM in the Uplink, 20+20 MHz	300.0	71.0

LTE is not only efficient for data but, because of a highly efficient uplink, is extremely efficient for VoIP traffic. As discussed in the "Spectral Efficiency" section above, in 10+10 MHz of spectrum, LTE VoIP capacity will reach 500 users.[156]

Table 19 analyzes LTE median and average throughput values in greater detail for different LTE configurations.

Table 19: LTE FDD User Throughputs Based on Simulation Analysis[157]

Configuration	User Throughput, Mbps			
	Downlink (DL)		Uplink (UL)	
	Median	Average	Median	Average
LTE FDD: Low Band, 2x2 MIMO-DL, 1x2 SIMO-UL, 10+10 MHz, R8	8.6	10.9	4.5	5.0
LTE FDD: High Band, 4x2 MIMO-DL, 1x4 SIMO-UL, 10+10 MHz, R8	10.6	12.2	5.4	6.4
LTE FDD: High Band, 2x2 MIMO-DL, 1x2 SIMO UL, 20+20 MHz, R8	15.2	17.9	5.4	7.0
LTE FDD: High Band, 4x4 MIMO-DL, 1x4 SIMO UL, 20+20 MHz, R12	25.4	29.2	6.9	9.1

---

[156] 3GPP Multi-member analysis.

[157] 5G Americas member contribution. SIMO refers to Single Input Multiple Output antenna configuration, which in the uplink means one transmit antenna at the UE and multiple receive antennas at the eNodeB.

The simulation results represent a consensus view of 5G Americas members working on this white paper project. The goal of the analysis was to quantify LTE throughputs in realistic deployments. Simulation assumptions include:

❑ Traffic is FTP-like at a 50% load with a 75/25 mix of indoor/outdoor users.

❑ Throughput is at the medium-access control (MAC) protocol layer. (Application-layer throughputs may be 5 to 8 percent lower due to protocol overhead.)

❑ The 3GPP specification release numbers shown correspond to the infrastructure capability.

❑ The configuration in the first row corresponds to low-frequency band operation, representative of 700 MHz or cellular, while the remaining configurations assume high-frequency band operation, representative of PCS, AWS, or WCS. (Higher frequencies facilitate higher-order MIMO configurations and have wider radio channels available.)

❑ The downlink value for the first row corresponds to Release 8 device-receive capability (Minimum Mean Square Error [MMSE]), while the values in the other rows correspond to Release 11 device-receive capability (MMSE – Interference Rejection Combining [IRC]).

❑ The uplink value for the first row corresponds to a Maximal Ratio Combining (MRC) receiver at the eNodeB, while the remaining values correspond to an IRC receiver.

❑ Low-band operation assumes 1,732-meter inter-site distance, while high-band operation assumes 500-meter ISD. The remaining simulation assumptions are listed in Table 20.

Table 20: LTE FDD User Throughput Simulation Assumptions[158]

Parameter	Value
Frequency	Low Band (LB): B17; High Band (HB): B30
Channel bandwidth	10 MHz, 20 MHz
System configuration	DL: 2x2, 4x2, and 4x4 Closed-Loop (CL) MIMO UL: 1x2 and 1x4 SIMO
Traffic type	FTP model 2: File size = 0.15 Mbyte, 1 second inter-arrival time, Load varied by changing number of users
Inter-Site Distance (ISD)	LB: 1732 m; HB: 500 m
Pathloss model	LB: HATA; HB: COST231 with correction
eNodeB transmit power	LB: 60 watts total; HB: 80 watts total
eNodeB antenna type	2 Tx = +-45 degrees cross-pol (DIV-1X); 4 Tx = Closely separated pair of cross-pols (CLA-2X)
eNodeB antenna gain	LB: 14.8 dBi; HB: 17.5 dBi
eNodeB antenna pattern	Actual antenna patterns as used in RF planning tool
eNodeB Rx type	LB: MRC; HB: IRC
Downtilt	LB: 7 degrees; HB: 9 degrees
Penetration loss	75/25 mix of indoor/outdoor users LB: 12 dB for indoor users; HB: 22 dB for indoor users
Device speed	3 km/h all users
Channel model	Modified SCME-WINNER+, LB: Suburban Macro (SMa) scenario; HB: Urban Macro (UMa)
Device antenna type	+-45 degrees cross-pol with built in correlation of 0.5
Device antenna gain and mismatch	LB: - 5 dBi and 3 dB; HB: -3 dBi and 3 dB
Device body loss	3 dB for both bands
Device Rx type	MMSE, MMSE-IRC
Uplink power control	LB: alpha = 1, Po = -100 dBm; HB: alpha = 0.9, Po = -100 dBm
Scheduler	Proportional fair, frequency selective

The assumptions, emphasizing realistic deployments, do not necessarily match assumptions used by other organizations, such as 3GPP, so results may differ.

Additional insight into LTE performance under different configuration comes from a test performed on a cluster of cells in an LTE operator's network, comparing downlink performance of 4X2 MIMO against 2X2 MIMO, and uplink performance of 1X4 SIMO against 1X2 SIMO. The test employed LTE category 4 devices.[159]

---

[158] 5G Americas member contribution.

[159] 5G Americas member contribution.

These tests, which were performed in a 20+20 MHz cluster, show significant improvements in cell edge uplink and downlink throughput, in addition to an overall increase in uplink and downlink throughputs. Specific results include:

❑ A 100% increase in uplink throughput at the cell edge with 1X4 SIMO compared to 1x2 SIMO.

❑ A 40% increase in downlink throughput at the cell edge with 4x2 closed-loop MIMO compared to 2x2 open-loop MIMO.

❑ A 50 to 75% increase in downlink throughput with closed loop MIMO compared to transmit diversity modes.

❑ Up to 6dB gains in uplink transmit power with 1X4 SIMO, which directly translates into UE battery savings.

❑ Peak speeds of 144 Mbps with 4X2 MIMO in the downlink and 47 Mbps with 1X4 SIMO in the uplink.

Another LTE operator's testing results for LTE in a TDD configuration, using 20 MHz channels, 3:2 DL to UL ratio, and category 3 devices, showed:

❑ Peak speeds of 55 Mbps.

❑ Typical speeds of 6 to 15 Mbps.[160]

Figure 74 shows the result of a drive test in a commercial LTE network with a 10 MHz downlink carrier demonstrating 20 Mbps to 50 Mbps throughput rates across much of the coverage area. Throughput rates would double with a 20+20 MHz configuration.

---

[160] 5G Americas member contribution.

Figure 74: Drive Test of Commercial European LTE Network (10+10 MHz)[161]

---

[161] Ericsson contribution.

Figure 75 provides additional insight into LTE downlink throughput, showing Layer 1 throughput simulated at 10 MHz bandwidth using the Extended Vehicular A 3 km/hour channel model. The figure shows the increased performance obtained with the addition of different orders of MIMO. Note how throughput improves based on higher signal to noise ratio (SNR).
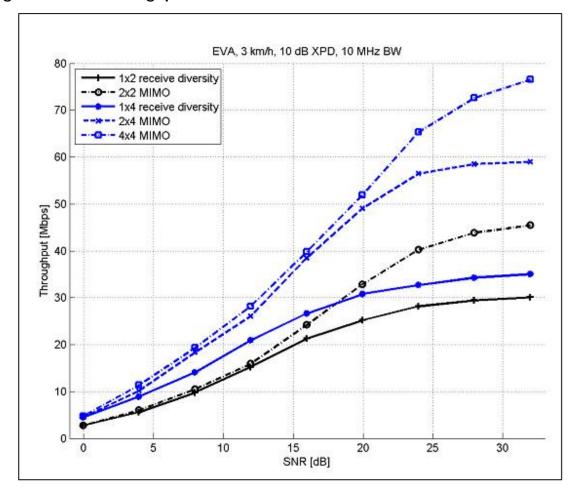
Figure 75: LTE Throughput in Various Modes[162]
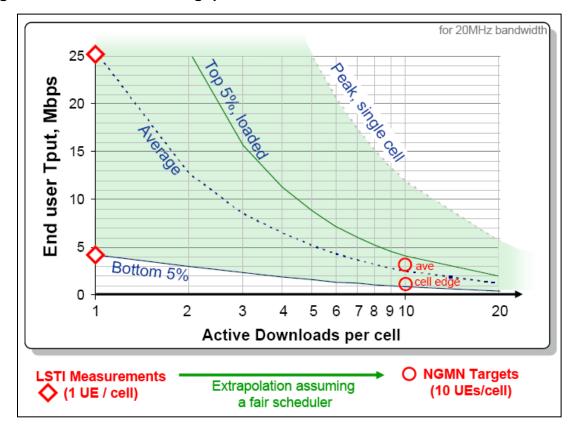
Actual throughput rates that users experience are lower than the peak rates and depend on a variety of factors:

❑ **RF Conditions and User Speed.** Peak rates depend on optimal conditions. Suboptimal conditions include being at the edge of the cell or moving at high speed, resulting in lower throughput.

❑ **Network Loading.** Like all wireless systems, throughput rates go down as more devices simultaneously use the network. Throughput degradation is linear.

---

[162] Jonas Karlsson, Mathias Refback, "Initial Field Performance Measurements of LTE," *Ericsson Review*, No. 3, 2008.

Figure 76 shows how dramatically throughput rates can vary by number of active users and radio conditions. The higher curves are for better radio conditions.

Figure 76: LTE Actual Throughput Rates Based on Conditions[163]

[163] LTE/SAE Trial Initiative, "Latest Results from the LSTI, Feb 2009," http://www.lstiforum.org.

VoLTE and RCS

This paper introduced VoLTE and voice support in the earlier section, "VoLTE, RCS, WebRTC, and Wi-Fi Calling." This section in the appendix provides additional technical detail about the operation of VoLTE and RCS.

Voice in LTE can encompass: no voice support, voice implemented in a circuit-switched fallback (CSFB) mode using 2G or 3G, and VoIP implemented with IMS.

Initial LTE network deployments used CSFB, with which the LTE network carries circuit-switched signaling over LTE interfaces, allowing the subscriber to be registered with the 2G/3G MSC even while on the LTE network. When there is a CS event, such as an incoming voice call, the MSC sends the page to the LTE core network, which delivers it to the subscriber device. The device then switches to 2G/3G operation to answer the call.

Voice over LTE using VoIP requires IMS infrastructure. To facilitate IMS-based voice, vendors and operators created the One Voice initiative to define required baseline functionality for user equipment, the LTE access network, the Evolved Packet Core, and the IMS. GSMA adopted the One Voice initiative in what it calls VoLTE, specified in GSMA reference document IR.92.[164] GSMA specifies interconnection and international roaming among LTE networks through the IR.88[165] specification. Another specification, IR.94, provides the IMS Profile for Conversational Video Service, a service referred to as "Video over LTE" (ViLTE).[166]

For a phone to support VoLTE, it needs software implementing the IMS protocol stack. For example, the iPhone 6 was the first iPhone to implement such software. Additional software implementing RCS application programming interfaces can provide applications with access to IMS-based services, such as voice, messaging, and video. The Open Mobile Alliance has defined RESTful network APIs for RCS that support the following functions: notification channel, chat, file transfer, third-party calls, call notification, video sharing, image sharing, and capability discovery. As shown in Figure 77, over time, new profile releases will broaden the scope of these APIs.

---

[164] GSMA, "IMS Profile for Voice and SMS," Document IR.92. Available at
http://www.gsma.com/newsroom/wp-content/uploads/2013/04/IR.92-v7.0.pdf.

[165] GSMA, "LTE Roaming Guidelines," GSMA Document IR.88. Available at
http://www.gsma.com/newsroom/wp-content/uploads/2013/04/IR.88-v9.0.pdf.

[166] GSMA, "IMS Profile for Conversational Video Service," Document IR.94. Available at
http://www.gsma.com/newsroom/all-documents/ir-94-ims-profile-for-conversational-video-service/.

Figure 77: Evolution of RCS API Profiles[167]

LTE VoIP leverages the QoS capabilities defined for EPC, which specify different quality classes. Features available in LTE to make voice operation more efficient include Semi-Persistent Scheduling (SPS) and TTI bundling. SPS reduces control channel overhead for applications (like VoIP) that require a persistent radio resource. Meanwhile, TTI bundling improves subframe utilization by reducing IP overhead, while in the process optimizing uplink coverage.

Another way to increase voice capacity in LTE and to support operation in congestion situations is vocoder rate adaptation, a mechanism with which operators can control the codec rate based on network load, thus dynamically trading off voice quality against capacity.

VoLTE roaming across operators will require network-to-network interfaces between their respective IMS networks. Such roaming and interconnect will follow initial VoLTE deployments. Different IMS stack implementations between vendors will also complicate roaming.

One roaming consideration is how operators handle data roaming. LTE roaming can send all visited network traffic back to the home network, which for a voice call, increases voice latency. For voice calls, the local breakout option would mitigate this latency.

Using Single-Radio Voice Call Continuity (SR-VCC) and Enhanced SR-VCC (eSRVCC), user equipment can switch mid-call to a circuit-switched network, in the event that the user moves out of LTE coverage. Similarly, data sessions can be handed over in what is called "Packet-Switched Handover" (PSHO).

---

[167] 4G Americas, *VoLTE and RCS Technology – Evolution and Ecosystem*, Nov. 2014.

Figure 78 shows how an LTE network might evolve in three stages. Initially, LTE performs only data service, and the underlying 2G/3G network provides voice service via CSFB. In the second stage, voice over LTE is available, but LTE covers only a portion of the total 2G/3G coverage area. Hence, voice in 2G/3G can occur via CSFB or SR-VCC. Eventually, LTE coverage will match 2G/3G coverage, and LTE devices will use only the LTE network.
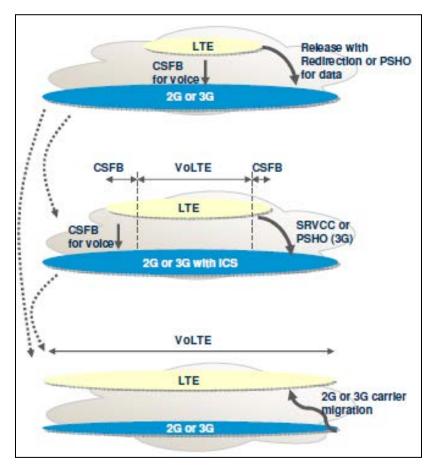
Figure 78: Evolution of Voice in an LTE Network[168]

Another voice approach, called "Voice over LTE via Generic Access" (VoLGA), defined circuit-switched operation through an LTE IP tunnel. 3GPP, however, has stopped official standards work that would support VoLGA.

3GPP has developed a new codec, called "Enhanced Voice Services" (EVS), which will include super-wideband voice capability. For the same bit rate, EVS provides higher voice quality than the other codecs.[169] Table 21 summarizes the features and parameters of the three 3GPP codecs used in LTE.

---

[168] 5G Americas member contribution.

[169] See Figure 9.2. 3GPP, *TR 26.952 V12.1.0, Codec for Enhanced Voice Services (EVS); Performance Characterization*, March 2015.

Table 21: Comparison of AMR, AMR-WB and EVS Codecs[170]

Features	AMR	AMR-WB	EVS
Input and output sampling frequencies supported	8KHz	16KHz	8KHz, 16KHz, 32KHz, 48 KHz
Audio bandwidth	Narrowband	Wideband	Narrowband, Wideband, Super-wideband, Fullband
Coding capabilities	Optimized for coding human voice signals	Optimized for coding human voice signals	Optimized for coding human voice and general-purpose audio (music, ringtones, mixed content) signals
Bit rates supported (in kb/s)	4.75, 5.15, 5.90, 6.70, 7.4, 7.95, 10.20, 12.20	6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85	5.9, 7.2, 8, 9.6 (NB and WB only), 13.2 (NB, WB and SWB), 16.4, 24.4, 32, 48, 64, 96, 128 (WB and SWB only)
Number of audio channels	Mono	Mono	Mono and Stereo
Frame size	20 ms	20 ms	20 ms
Algorithmic Delay	20-25 ms	25 ms	Up to 32 ms

Figure 79 shows mean opinion scores (MOS) for different codecs at different bit rates, illustrating the advantage of EVS, particularly for bit rates below 32 kbps that cellular networks use.

---

[170] 4G Americas, *Mobile Broadband Evolution: Rel-12 & Rel-13 and Beyond*, 2015. See also T-Mobile 2016 EVS announcement: https://newsroom.t-mobile.com/news-and-blogs/volte-enhanced-voice-services.htm.
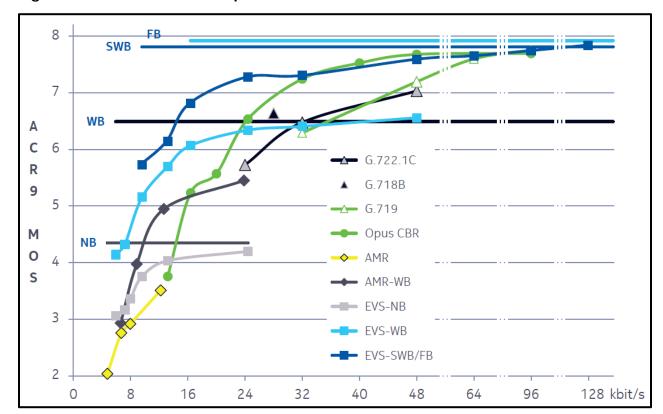
Figure 79: Combined Mean Opinion Score Values[171]

Table 22 shows EVS (narrowband, wideband, super-wideband) audio bandwidths and bitrates that create subjective quality equal to or better than AMR or AMR-WB for typical conversational voice scenarios.

Table 22: EVS Compared to AMR and AMR-WB[172]

Reference	Equal bandwidth	Wider bandwidth
AMR 12.2 kbit/s	EVS-NB 8.0 kbit/s	EVS-WB 5.9 kbit/s
AMR-WB 12.65 kbit/s	EVS-WB 9.6 kbit/s	EVS-SWB 9.6 kbit/s
AMR-WB 23.85 kbits/s	EVS-WB 13.2 kbit/s	EVS-SWB 9.6 kbit/s

Figure 80 compares EVS capacity gains over AMR and AMR-WB for the references cases shown in Table 22. EVS-SWB at 9.6 kbps almost doubles voice capacity compared to AMW-WB at 23.85 kbps.

---

[171] Nokia, *The 3GPP Enhanced Voice Services (EVS) codec*, 2015.

[172] Ibid.
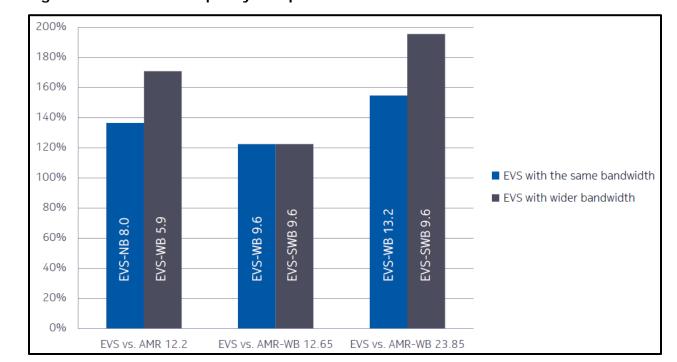
Figure 80: EVS Voice Capacity Compared to AMR and AMR-WB[173]

LTE Ultra-Reliable and Low-Latency Communications

The 3GPP work item for this feature states, "3GPP LTE technology enhancements are needed to serve such new use cases and to remain technologically competitive up to and beyond 2020. As a candidate technology for ITU IMT-2020 submission, it is motivated to further enhance the LTE system such that it can meet the key IMT-2020 requirements including those for URLLC in terms of reliability ($1 \text{-} 10^{-5}$ reliability for small data packets within a latency of 1ms) as well as latency (≤1ms one way user plane latency)."[174]

Evolved Packet Core (EPC)

3GPP defined the Evolved Packet Core (EPC) in Release 8 as a framework for an evolution or migration of the network to a higher-data-rate, lower latency, packet-optimized system that supports multiple radio-access technologies including LTE, as well as and legacy GSM/EDGE and UMTS/HSPA networks. EPC also integrates CDMA2000 networks and Wi-Fi.

EPC is optimized for all services to be delivered via IP in a manner that is as efficient as possible—through minimization of latency within the system, for example. It also provides service continuity across heterogeneous networks, which is important for LTE operators who must simultaneously support GSM-HSPA customers.

One important performance-enhancing aspect of EPC is a flatter architecture. For packet flow, EPC includes two network elements, called "Evolved Node B" (eNodeB) and the Access Gateway (AGW). The eNodeB (base station) integrates the functions traditionally

---

[173] Ibid.

[174] RP-170796, 3GPP Work Item Description, "Ultra Reliable Low Latency Communication for LTE," March 2017.

performed by the radio network controller, which previously was a separate node controlling multiple Node Bs. Meanwhile, the AGW integrates the functions traditionally performed by the SGSN and GGSN. The AGW includes both control functions, handled through the Mobile Management Entity (MME), and user plane (data communications) functions. The user plane functions consist of two elements: A serving gateway that addresses 3GPP mobility and terminates eNodeB connections, and a Packet Data Network (PDN) gateway that addresses service requirements and also terminates access by non-3GPP networks. The MME serving gateway and PDN gateways can be collocated in the same physical node or distributed, based on vendor implementations and deployment scenarios.

EPC uses IMS as a component. It also manages QoS across the whole system, an important enabler for voice and other multimedia-based services.

Figure 81 shows the EPC architecture.

Figure 81: EPC Architecture

Elements of the EPC architecture include:

❑ Support for legacy GERAN and UTRAN networks connected via SGSN.

❑ Support for new radio-access networks such as LTE.

❑ Support for non-3GPP networks such as EV-DO and Wi-Fi. (See section below on Wi-Fi integration).

❑ The Serving Gateway that terminates the interface toward the 3GPP radio-access networks.

- The PDN gateway that controls IP data services, does routing, allocates IP addresses, enforces policy, and provides access for non-3GPP access networks.

- The MME that supports user equipment context and identity, as well as authenticating and authorizing users.

- The Policy Control and Charging Rules Function that manages QoS aspects.

QoS in EPS employs the QoS Class Identifier (QCI), a number denoting a set of transport characteristics (bearer with/without guaranteed bit rate, priority, packet delay budget, packet error loss rate) and used to infer nodes specific parameters that control packet forwarding treatment (such as scheduling weights, admission thresholds, queue management thresholds, or link-layer protocol configuration). The network maps each packet flow to a single QCI value (nine are defined in the Release 8 version of the specification) according to the level of service required by the application. Use of the QCI avoids the transmission of a full set of QoS-related parameters over the network interfaces and reduces the complexity of QoS negotiation. The QCI, together with Allocation Retention Priority (ARP) and, if applicable, Guaranteed Bit Rate (GBR) and Maximum Bit Rate (MBR), determines the QoS associated to an EPS bearer. A mapping between EPS and pre-Release 8 QoS parameters permits interworking with legacy networks.

The QoS architecture in EPC enables a number of important capabilities for both operators and users:

- **VoIP support with IMS.** QoS is a crucial element for providing LTE/IMS voice service. (See section below on IMS).

- **Enhanced application performance.** Applications such as gaming or video can operate more reliably.

- **More flexible business models.** With flexible, policy-based charging control, operators and third parties will be able to offer content in creative new ways. For example, an enhanced video stream to a user could be paid for by an advertiser.

- **Congestion control.** In congestion situations, certain traffic flows (bulk transfers, abusive users) can be throttled down to provide a better user experience for others.

Table 23 shows the nine QCI used by LTE.

Table 23: LTE Quality of Service

QCI	Resource Type	Priority	Delay Budget	Packet Loss	Examples
1	GBR (Guaranteed Bit Rate)	2	100 msec.	10^{-2}	Conversational voice
2	GBR	4	150 msec.	10^{-3}	Conversational video (live streaming)
3	GBR	3	50 msec.	10^{-3}	Real-time gaming
4	GBR	5	300 msec.	10^{-6}	Non-conversational video (buffered streaming)
5	Non-GBR	1	100 msec.	10^{-6}	IMS signaling

QCI	Resource Type	Priority	Delay Budget	Packet Loss	Examples
6	Non-GBR	6	300 msec.	10^{-6}	Video (buffered streaming), TCP Web, email, and FTP
7	Non-GBR	7	100 msec.	10^{-3}	Voice, video (live streaming), interactive gaming
8	Non-GBR	8	300 msec.	10^{-6}	Premium bearer for video (buffered streaming), TCP Web, e-mail, and FTP
9	Non-GBR	9	300 msec.	10^{-6}	Default bearer for video, TCP for non-privileged users

Heterogeneous Networks and Small Cells

A fundamental concept in the evolution of next-generation networks is the blending of multiple types of networks to create a "network of networks" characterized by:

- ❑ Variations in coverage areas, including femtocells (either enterprise femtos or home femtos, called HeNBs), picocells (also referred to as metro cells), and macro cells. Cell range can vary from 10 meters to 50 kilometers.

- ❑ Different frequency bands.

- ❑ Different technologies spanning Wi-Fi, 2G, 3G, 4G, and 5G.

- ❑ Relaying capability in which wireless links can serve as backhaul.

Figure 82 shows how user equipment might access different network layers.
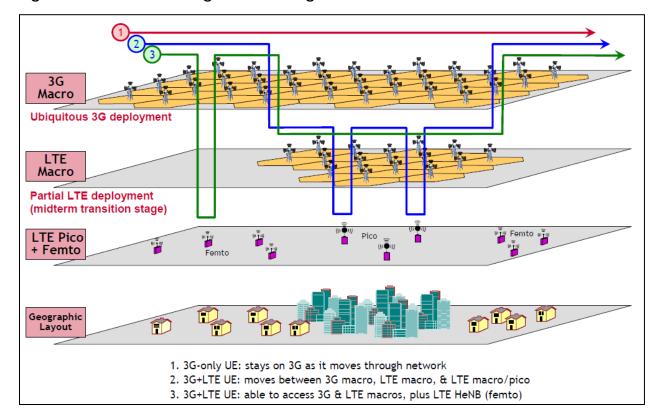
Figure 82: Load Balancing with Heterogeneous Networks[175]

HetNets will allow significant capacity expansion in configurations in which operators can add picocells to coverage areas served by macrocells, particularly if there are hot spots with higher user densities.

Small cells differentiate themselves from macrocells according to the parameters shown in Table 24.

Table 24: Small Cell Vs. Macro Cell Parameters: Typical Values

Parameter	Small Cell	Macro Cell
Transmission Power	24 dBm (0.25 W)	43 dBm (20 W)
Antenna Gain	2 dBi	15 dBi
Users	Tens	Hundreds
Mobility	30 km/hr	350 km/hr

---

[175] 5G Americas member contribution.

Whether or not the small cell uses the same radio carriers as the macro cell involves multiple tradeoffs. In Figure 83 Scenario 1, the small cells and macro cell use different radio carriers, the two not interfering with each other. Although this configuration requires more spectrum, the small cells are able to cover larger areas than if they were deployed using the same radio carrier as the macro. This configuration supports medium-to-high penetration levels of small cells, allowing the network to reach huge capacity.

In Scenario 2, the small cells and macro cells use the same radio carrier, accommodating operators with more limited spectrum, but the network must manage interference using the techniques discussed below. Operators must carefully manage small-cell transmission power in this configuration.

Figure 83: Scenarios for Radio Carriers in Small Cells

In Scenario 3, the small cells use a straddled radio carrier, accommodating operators with more spectrum, but the network still needs to manage interference using techniques discussed below. Compared with a shared carrier configuration, this configuration has benefits similar to dedicated carriers in terms of radio-parameter planning and reduced interference.

Figure 84 shows two different traffic distribution scenarios, with a uniform distribution of devices in the first and higher densities serviced by picocells in the second. The second scenario can result in significant capacity gains as well as improved user throughput.
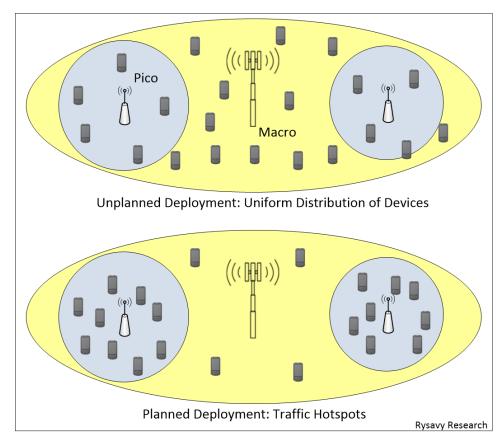
Figure 84: Different Traffic Distributions Scenarios

Unplanned Deployment: Uniform Distribution of Devices

Planned Deployment: Traffic Hotspots

Rysavy Research

One vendor calculated expected HetNet gains assuming no eICIC, no picocell range extension, and no eICIC. For the case of four picocells without picocell range extension and uniform user distribution, the median-user-throughput gain compared with a macro-only configuration was 85%. For a similar case of four picocells but using a hotspot user distribution, the gain was much higher, 467%.[176] Additional gains will occur with picocell range extension.

Expected picocell gains rise proportionally to the number of picocells, so long as a sufficient number of UEs connect to the picocells.

Release 10 and Release 11 added enhanced support to manage the interference in the HetNet scenario in the time domain with Enhanced Inter-cell Interference Coordination (eICIC) and Further Enhanced Intercell Interference Coordination (feICIC), as well as in the frequency domain with carrier-aggregation-based ICIC.

HetNet capability keeps becoming more sophisticated through successive 3GPP releases as summarized in Table 25.

---

[176] 5G Americas member contribution. Further assumes 2X1 W picocell transmit power, cell-edge placement (planned picocell deployment), 67% of all the users within 40m of the pico locations, and 3GPP Technical Report 36.814 adapted to 700 MHz.

Table 25: 3GPP HetNet Evolution

3GPP Release	HetNet Feature
8	Initial SON capabilities, most for auto configuration. Initial intercell interference coordination (ICIC) available.
9	More mobility options (for example, handover between HeNBs), operator customer subscriber group (SCG) lists, load-balancing, coverage and capacity improvements.
10	An interface for HeNBs, called "Iurh," that improves coordination and synchronization, LTE time domain eICIC. Carrier-aggregation-based ICIC also defined.
11	Improved eICIC, further mobility enhancements.

Enhanced Intercell Interference Coordination

Significant challenges must be addressed in these heterogeneous networks. One is near-far effects, in which local small-cell signals can easily interfere with macro cells if they are using the same radio carriers.

Interference management is of particular concern in HetNets since, by design, coverage areas of small coverage cells overlap with the macro cell. Beginning with Release 10, eICIC introduces an approach of almost-blank subframes by which subframe transmission can be muted to prevent interference. Figure 85 illustrates eICIC for the macro layer and pico layer coordination. If a UE is on a picocell but in a location where it is sensitive to interference from the macro layer, the macro layer can mute its transmission during specific frames when the pico layer is transmitting.
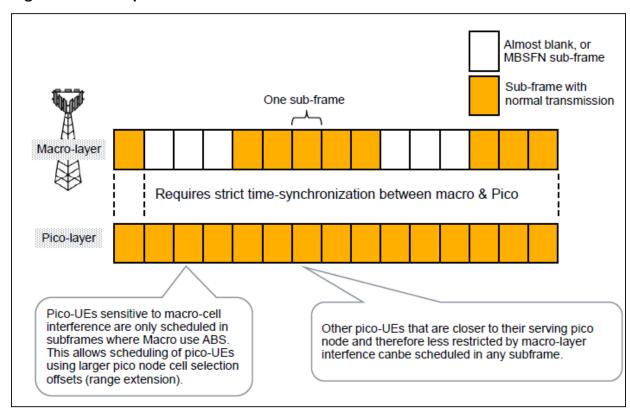
Figure 85: Example of Enhanced Intercell Interference Coordination[177]

LTE can also combine eICIC with interference-cancellation-based devices to minimize the harmful effects of interference between picocells and macro cells.

Figure 86 shows one 4G America member's analysis of anticipated median throughput gains using picocells and Release 11 Further Enhanced ICIC.

---

[177] 5G Americas member contribution.
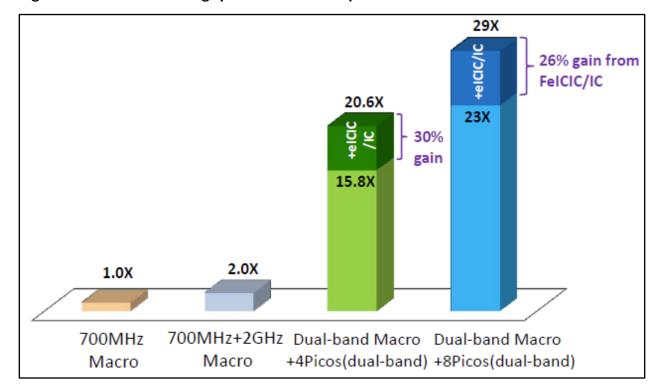
Figure 86: Median Throughput Gains in Hotspot Scenarios[178]
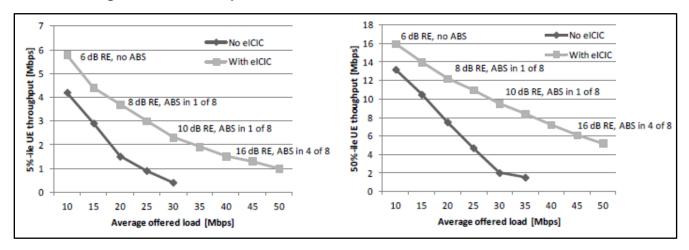
FeICIC is also beneficial in non-hotspot scenarios. In the case of a uniform distribution of picocells, this same 5G Americas member estimates a 130% gain from FeICIC for an eight picocell per macro-cell scenario, increasing capacity from a factor of 3.3 for the picocells alone to a factor of 7.6 with the addition of FeICIC.[179]

Further insight is available from Figure 87, which shows 5 percentile and 50 percentile throughput with and without eICIC under different conditions of range extension and almost blanked subframes.

---

[178] 5G Americas member contribution. Assumes 3GPP evaluation methodology TR 36.814, carrier-aggregation UEs, macro ISD = 1732m, 700 MHz and 2GHz carrier frequency, full-buffer traffic, FDD 10+10 MHz per carrier, 6-degree antenna downtilt, 4 or 8 Picos and 30 UEs per Macro cell, hotspot distribution with 20 of 30 UEs near picos, PF scheduler, 2x2 MIMO, TU3 channel, NLOS, local partitioning algorithm.

[179] Assumes 3GPP evaluation methodology TR 36.814, macro ISD = 1732m, 700 MHz and 2GHz carrier frequency, full-buffer traffic, 6-degree antenna downtilt, 30 carrier-aggregation UEs per Macro cell, uniform random layout, PF scheduler, FDD, 10+10 MHz per carrier, 2x2 MIMO, TU3 channel, NLOS, local partitioning algorithm. Additional information is available at ftp://ftp.3gpp.org/tsg_ran/WG1_RL1/TSGR1_66b/Docs/R1-113383.zip.

Figure 87: User Throughput Performance With/Without eICIC for Dynamic Traffic vs. Average Offered Load per Macro Cell Area[180]

The muting of certain subframes in eICIC is dynamic and depends on identifying, on a per user basis, whether an interfering cell's signal exceeds a threshold relative to the serving cell signal. Coordinating muting among small cells can be complicated because a small cell can simultaneously be an interferer while serving a UE that is a victim of another cell. The network must therefore coordinate muting among multiple small cells.

Figure 88 below at left shows user throughput gains of time domain interference relative to network load. Throughput gains are higher at higher network loads because of more active users and the higher likelihood of interference between the small cells.

Figure 88 below at right shows the maximum muting ratio, which increases with higher network load.

---

[180] 5G Americas member contribution. Assumes 3GPP evaluation methodology TR 36.814, 500 meter ISD, 4 picos per macro-cell area, Poisson call arrival, finite payload for each call, and termination of call upon successful delivery.
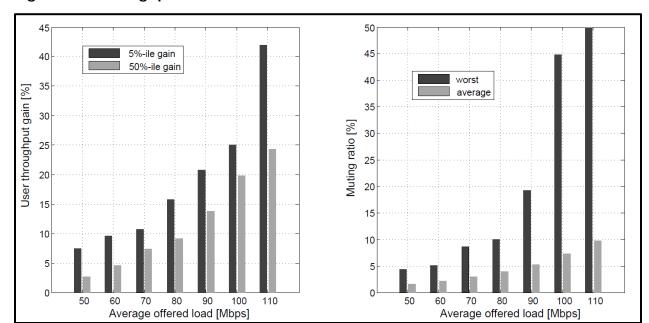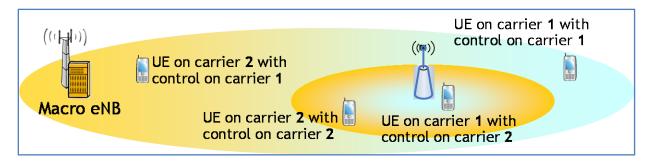
Figure 88: Throughput Gain of Time-Domain Interference Coordination[181]

Another approach for addressing inter-layer interference cancellation in HetNets can come from carrier aggregation with no further additions or requirements and realizable with Release 10 LTE networks. Consider the scenario in Figure 89, in which both the macro eNB and the pico eNB are allocated two component carriers (namely CC1 and CC2). The idea is to create a "protected" component carrier for downlink control signals and critical information (Physical Downlink Control Channel, system information, and other control channels) while data can be conveniently scheduled on both component carriers through cross-carrier scheduling.

Figure 89: Carrier-Aggregation Based ICIC[182]

CC1 is the primary component carrier for the macro cell, while CC2 is the primary for the picocell; hence the protected carriers are CC1 for the macro cell and CC2 for the picocell. The macro cell allocates a lower transmission power for its secondary CC in order to reduce

---

[181] 5G Americas member contribution. Simulations based on 12 densely deployed small cells at 3.5 GHz and 3GPP Release 12 simulation assumptions in TR 36.842.

[182] 5G Americas member contribution.

interference to the picocell's primary component carrier. The network can schedule data on both the primary and secondary component carriers. In the figure, users in the cell range expansion (CRE) zone can receive data via cross-carrier scheduling from the secondary CC at subcarrier frequencies on which interference from the other cell can be reduced if the cells exchange appropriate signaling over what is called an "X2 interface." Users operating close to the eNodeBs can receive data from both component carriers as their interference levels will hopefully be lower. Therefore, a CA-capable receiver will enjoy the enhanced throughput capabilities of carrier aggregation, while simultaneously receiving extra protection for control and data channels at locations with potentially high inter-layer interference.

Thus, carrier aggregation can be a useful tool for deployment of heterogeneous networks without causing a loss of bandwidth. These solutions, however, do not scale well (in Release 10 systems) to small system bandwidths (say, 3+3 MHz or 1.4+1.4 MHz radio carriers) because control channels occupy a high percentage of total traffic. Additionally, interference between the cell reference signals (CRS) would also be significant.

Dual Connectivity

A major enhancement in Release 12 is a UE being served at the same time by both a macro cell and a small cell operating at different carrier frequencies, a capability called dual connectivity and illustrated in Figure 90. Data first reaches the macro eNodeB and is split, with part of it transmitted from the macro and the balance sent via an X2 interface to the small cell for transmission to the UE.

Figure 90: Dual Connectivity[183]
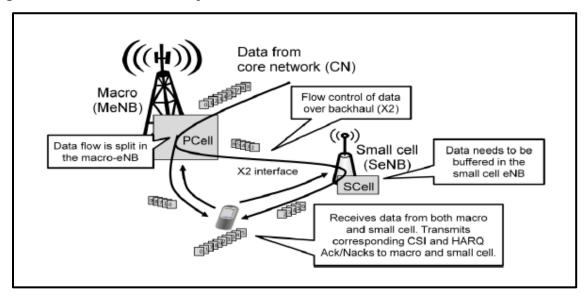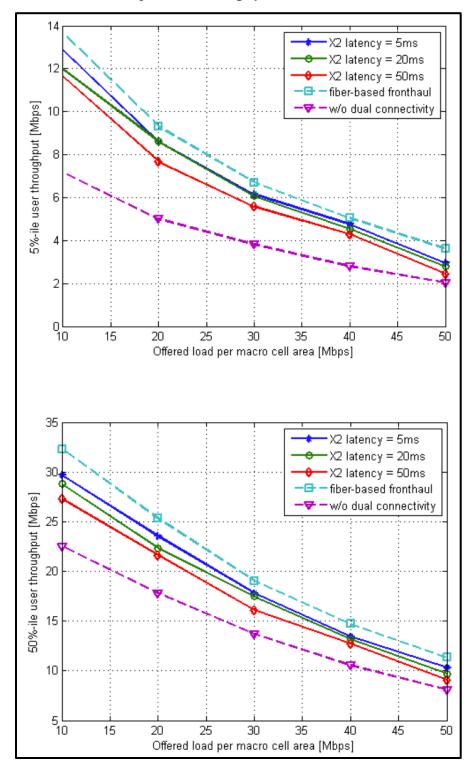
Figure 91 shows throughput gains of dual connectivity at 5 percentile and 50 percentile (median) levels relative to the load on the network and different degrees of latency in the X2 interface. Benefits are higher with lower network load and with lower X2 latency.

---

[183] Source: 5G Americas member contribution.

Figure 91: Dual Connectivity User Throughput[184]

---

[184] 5G Americas member contribution.

Internet of Things and Machine-to-Machine

Anticipating huge growth in machine-to-machine communications, Release 11 added a Machine Type Communications (MTC) Interworking Function and Service Capability Server. Release 12 defined a category 0 device designed to deliver low cost through a single antenna design and other simplifications.[185] Release 13 went even further, with a category M-1 architecture that further reduces cost, improves range, and extends battery life. Category 13 also added Narrowband-IoT capability with Category NB-1 and an IoT solution for GSM, called "EC-GSM-IoT," that extends coverage by 20 dB. Category M-1 and NB-IoT devices could achieve battery life as high as 10 years.

Figure 92 depicts the methods used to reduce cost in a Category M device compared with a Category 4 device.
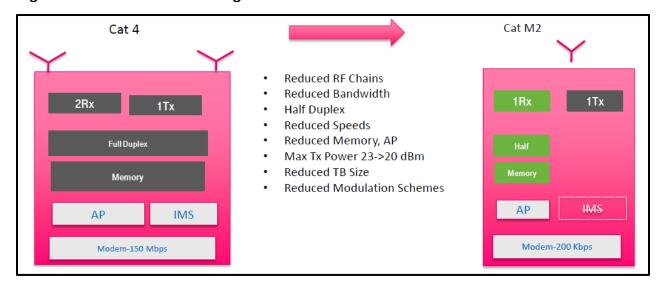
Figure 92: Means of Achieving Lower Cost in IoT Devices[186]

Table 26 summarizes the features of different LTE IoT devices based on 3GPP Release.

Table 26: Summary of IoT Features in LTE Devices

Device Category	Category 3	Category 1	Category 0	Category M-1	Category NB-1	EC-GSM-IoT
3GPP Release	10	11	12	13	13	13

---

[185] 3GPP, *Access System for Ultra Low Complexity and Low Throughput Internet of Things based on Cellular,* GP-140301, May 2014.

[186] 5G Americas member contribution.

Device Category	Category 3	Category 1	Category 0	Category M-1	Category NB-1	EC-GSM-IoT
Max. Data Rate Downlink	100 Mbps	10 Mbps	1 Mbps	1 Mbps	200 Kbps	74 Kbps
Max. Data Rate Uplink	50 Mbps	5 Mbps	1 Mbps	1 Mbps	200 Kbps	74 Kbps
Max. Bandwidth	20 MHz	20 MHz	20 MHz	1.08 MHz	0.18 MHz	0.2 MHz
Duplex	Full	Full	Optional half-duplex	Optional half-duplex	Half	Half
Max. Receive Antennas	Two	Two	One	One	One	One
Power		Power Save Mode[187]	Power Save Mode	Power Save Mode		
Sleep				Longer sleep cycles using Idle Discontinuous Reception (DRX)		
Coverage				Extended through redundant transmissions and Single Frequency Multicast		

Cloud Radio-Access Network (RAN) and Network Virtualization

Still in the early stages of development, cloud RAN (C-RAN) is a distributed architecture in which multiple remote radio heads connect to a "cloud" that consists of a farm of baseband processing nodes. This approach can improve centralized processing, as is needed for

---

[187] Power Save Mode specified in Release 12, but applicable to Category 1 device configured as Release 12.

CoMP, centralized scheduling, and Multiflow, without the need to exchange information among many access nodes. The performance of both LTE and HSPA technologies could be enhanced by the application of cloud RAN architectures. The term "fronthauling" has been used to describe the transport of "raw" radio signals to central processing locations, such as between the Physical Network Function (PNF) and a Virtual Network Function (VNF).

This architecture, shown in Figure 93, comes at the cost of requiring high-speed, low-latency backhaul links between these radio heads and the central controller. One vendor states that carrying 10+10 MHz of LTE with 2X2 MIMO requires 2.5 Gbps of bandwidth and imposes less than 0.1 msec of delay.[188] A standard called "Common Public Radio Interface" (CPRI) addresses generic formats and protocols for such a high-speed link. ETSI has also developed the Open Radio Equipment Interface (ORI). The feasibility of cloud RAN depends to a large extent on the cost and availability of fiber links between the remote radio heads and the centralized baseband processing location.

Unlike virtualizing the EPC, in which the entirety of the function can be virtualized, cloud RAN needs a PNF that terminates the RF interface. Cloud RAN therefore requires a split to be defined within the RAN. As a consequence, initial deployments of cloud RAN have looked to ruse the CPRI interface between the RRH and the baseband unit.
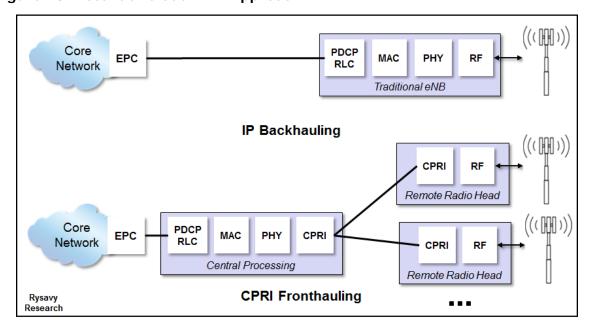
Figure 93: Potential Cloud RAN Approach

The next evolutionary step after centralizing baseband processing is to virtualize the processing by implementing the functions in software on commodity computing platforms, thus abstracting the functions from any specific hardware implementation.

C-RANs can vary by the extent of coverage, ranging from being highly localized and operating across a small number of sites to metropolitan-wide solutions. Other variables

---

[188] Dudu Bercovich, Ran Avital, "Holistic HetNet Hauling (3H)," Ceragon, February 2013. Available at http://www.ceragon.com/images/Reasource_Center/White_Papers/Ceragon_Holistic_Hetnet_Hauling_White_Paper.pdf.

include existing deployments versus greenfield situations, new LTE and 5G technologies versus integrating legacy 2G and 3G technologies, and integrating Wi-Fi. Greater scope increases complexity but yields benefits including better load-balancing and greater flexibility in spectrum re-farming.

Another design choice, as detailed in Table 27, is whether to centralize Layer 1 and Layer 2 functions (an RF-PHY split), or whether to keep Layer 1 at the base stations and centralize only Layer 2 (a PHY-MAC split).

Table 27: Partially Centralized Versus Fully Centralized C-RAN

	Fully Centralized	**Partially Centralized**
Transport Requirements	Multi-Gbps, usually using fiber	20 to 50 times less
Fronthaul Latency Requirement	Less than 100 microseconds	Greater than 5 milliseconds
Applications	Supports eICIC and CoMP	Supports centralized scheduling
Complexity	High	Lower
Benefit	Capacity gain	Lower capacity gain

Figure 94 analyzes the different possible RAN decompositions in greater detail.

Figure 94: Costs and Benefits of Various RAN Decompositions[189]

Legend:
- No additional cost of benefit enabled
- Cost added or challenge to providing benefit
- Major cost added or challenge to providing benefit

	Complete Waveform Shipped to RF (IQ Samples)	Waveform Modulation/Demod distributed	Distribute PHY and centralize MAC	Distribute time sensitive MAC (e.g., HARQ process)	Distribute Radio Link Control to reduce time sensitivity	Control Pane/Data Plane Split
	CPRI	Split PHY	MAC-PHY	Split MAC	PDCP-RLC	RRC-PDCP Split
Fronthaul delay requirements	100 us transport latency	<6 ms latency for interleaved HARQ	<6 ms latency for interleaved HARQ	RLC ACK Windowing latency only	Same as legacy backhaul	Same as legacy backhaul
Fronthaul bandwidth requirements	30 x BW expansion	UL BW expansion due to soft bits	Same as legacy backhaul	Same as legacy backhaul	Same as legacy backhaul	Same as legacy backhaul
Multi-vendor alignment	Limited multi-vendor ORI Support	Proprietary	Small Cell Ecosystem defining	Proprietary	Challenging	Challenging
Virtualization Support	Specialized HW required	Some functions virtualized	Some functions virtualized	Virtualized central functions	Maximal virtualization	Maximal virtualization
Performance Improvements	Inter-cell gains possible	Inter-cell gains possible	Some inter-cell gains possible	Limited inter-cells gains	Limited inter-cells gains	Some handover optimization

Although some operators in dense deployments with rich fiber assets may centralize all functions, Figure 94 uses the red rectangles to show the two most likely functional splits for LTE-Advanced and 5G:

1. **Distributed PHY and Centralized MAC.** This approach relaxes the fronthaul delay requirement to 6 msec, compared with the CPRI requirement of 250 microseconds. Fronthaul bandwidth requirement is only 10-20% greater than conventional backhaul.

2. **Control Plane/Data Plane Split.** This approach further relaxes fronthaul requirements to 30 msec and is the approach used for dual-connectivity, such as a macro and small cell simultaneously connecting to a user. For 5G, 3GPP Release 15 specifications standardize a split for cloud RAN between the PDCP and RLC layers.

---

[189] Cisco, *Cisco 5G Vision Series: Small Cell Evolution*, 2016.

Next Generation Mobile Networks studied the pros and cons of different fronthauling interfaces and published the results in March 2015.[190]

Longer-term, perhaps in the 5G context, virtualized C-RANs may take away the very concept of cells. With methods such as beamforming and device-to-device communication, coverage may extend dynamically from a multitude of sources based on instantaneous load notifications and the radio resources available at different nodes.

In the past, RAN and core networks have been distinct entities, but over the next decade, the two may merge with more centralized, virtualized, and cloud-driven approaches.

Another form of virtualization is software-defined networking, an emerging trend in both wired and wireless networks. For cellular, SDN promises to reduce OPEX costs, simplify the introduction of new services, and improve scalability; all major infrastructure vendors are involved. The Open Networking Foundation explains that an SDN decouples the control and data planes, centralizing network state and intelligence, while abstracting the underlying network infrastructure from applications.[191] Virtualization of network functions will be a complex, multi-year undertaking and will occur in stages, as shown in Figure 95.

---

[190] Next Generation Mobile Networks, *Further Study on Critical C-RAN Technologies, Version 1.0*, March 2015. See sections 2.2 and 2.3. Available at https://www.ngmn.org/uploads/media/NGMN_RANEV_D2_Further_Study_on_Critical_C-RAN_Technologes_v1.0.pdf.

[191] Open Networking Foundation, "Software-Defined Networking: The New Norm for Networks," http://www.opennetworking.org/sdn-resources/sdn-library/whitepapers, accessed June 20, 2014.
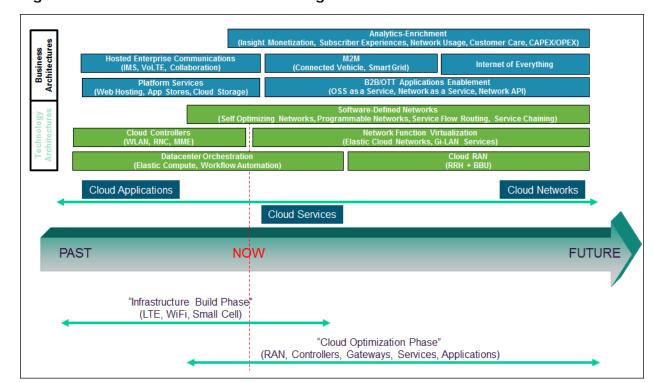
Figure 95: Software-Defined Networking and Cloud Architectures[192]

Unlicensed Spectrum Integration

See the earlier section in this report on unlicensed spectrum integration, which includes a discussion of LTE-U, LTE-LAA, MulteFire, LWA, LWIP, and RCLWI. This section covers integration approaches other than these.

3GPP has evolved its thinking on how best to integrate Wi-Fi with 3GPP networks. At the same time, the Wi-Fi Alliance and other groups have also addressed hotspot roaming, namely the ability to enable an account with one public Wi-Fi network provider to use the services of another provider that has a roaming arrangement with the first provider.

The multiple attempts to make Wi-Fi networks universally available have made for a confusing landscape of integration methods, which this section attempts to clarify. Most integration today is fairly loose, meaning that either a device communicates data via the cellular connection or via Wi-Fi. If via Wi-Fi, the connection is directly to the internet and bypasses the operator core network. In addition, any automatic handover to hotspots occurs only between the operator cellular network and operator-controlled hotspots. The goals moving forward are to:

❑ Support roaming relationships so that users can automatically access Wi-Fi hotspots operated by other entities.

❑ Enable automatic connections so that users do not have to enter usernames and passwords. In most cases, this will mean authentication based on SIM credentials.

---

[192] 5G Americas member contribution.

- ❑ Provide secure communications on the radio link as provided by the IEEE 802.11i standard.

- ❑ Allow policy-based mechanisms that define the rules by which devices connect to various Wi-Fi networks.

- ❑ Enable simultaneous connections to both cellular and Wi-Fi, with control over which applications use which connections.

- ❑ Support different types of Wi-Fi deployments, including third-party access points and carrier access points.

Release 6 I-WLAN

3GPP Release 6 was the first release to offer the option of integrating Wi-Fi in a feature called "Interworking WLAN" (I-WLAN), using a separate IP address for each network type.

Release 8 Dual Stack Mobile IPv6 and Proxy Mobile IPv6

3GPP Release 8 specified Wi-Fi integration with the EPC using two different approaches: host-based mobility with Dual Stack Mobile IPv6 (DSMIPv6) in the client, and network based mobility with Proxy Mobile IPv6 (PMIPv6) using an intermediary node called an "Enhanced Packet Data Gateway" (ePDG).[193] This method is intended for untrusted (non-carrier-controlled) Wi-Fi networks.

Release 11 S2a-based Mobility over GTP

Release 11, however, implements a new and advantageous approach as shown in Figure 96, one that eliminates the ePDG. Called "S2a-based Mobility over GTP" (SaMOG), a trusted WLAN Access Gateway connects to multiple 3GPP-compliant access points. Traffic can route directly to the internet or traverse the packet core. This method is intended for trusted (carrier-controlled) Wi-Fi networks.

---

[193] 3GPP, *System Architecture Evolution (SAE); Security aspects of non-3GPP accesses.* TS 33.402.

Figure 96: Release 11 SaMOG-based Wi-Fi Integration

Release 12 improves SaMOG capabilities in Enhanced SaMOG (eSaMOG), in which UEs can:

❑ Request the connectivity type

❑ Indicate the Access Point Name (APN) to establish PDN connectivity

❑ Request to hand over an existing PDN connection

❑ Establish multiple PDN connections in parallel over the WLAN

❑ Establish a non-seamless WLAN offload connection in parallel to a Packet Data Network connection over WLAN.

Multipath TCP

A new method for potentially integrating Wi-Fi and 3GPP networks is based on work by the Internet Engineering Taskforce (IETF). Called "Multipath TCP," the approach allows a TCP connection to occur simultaneously over two different paths. The advantages of this approach include higher speeds by aggregating links and not requiring any special provisions for link-layer handovers.

The IETF has published an experimental specification, *Request for Comments 6824: CP Extensions for Multipath Operation with Multiple Addresses*, which explains this approach. The IETF is also specifying Multipath QUIC.

ANDSF

Another relevant specification is 3GPP Access Network Discovery and Selection Function (ANDSF), which provides mechanisms by which mobile devices can know where, when, and how to connect to non-3GPP access networks, such as Wi-Fi.[194] ANDSF operates independently of SaMOG or other ways that Wi-Fi networks might be connected.

---

[194] 3GPP, *Architecture enhancements for non-3GPP accesses, Technical Specification 23.402*.

ANDSF functionality increases with successive 3GPP versions, as summarized in Table 28.

Table 28: ANDSF Policy Management Objects and 3GPP Releases[195]

ANDSF Policy Type	Policy Rule & Management Object	Release 8, 9	Release 10, 11	Release 12
Inter-System Mobility Policy (ISMP)	Policy, Rule priority, Prioritized Access, Validity Area (3G,4G, WI-FI, Geo), PLMN, Time-of-Day	X	X	X
Discovery Info	Access Network Type, Access Network Area (3G, 4G, Wi-Fi, Geo), Access Network Reference	X	X	X
UE Location	3GPP, 3GPP2, WiMAX, Wi-Fi network ID, Geo Location, PLMN	X	X	X
Inter-System Routing Policy (ISRP)	Flow Based routing, Service Based routing, Non-Seamless Offload, Roaming, PLMN, Routing Criteria, Time-of-Day, Routing rule		X	X
UE Profile	Device app/OS capability		X	X
Inter-APN Routing Policy (IARP)	Inter-APN routing over IP interface (in progress)			X
WLAN Selection Policy	Operator defined WLAN selection policy			X
Rule Selection Information	VPLMN with preferred WLAN roaming			X
Home Operator Preference	Home SP preference for S2a PDN session			X

Bidirectional Offloading Challenges

Eventually, operators will be able to closely manage user mobile broadband and Wi-Fi connections, dynamically selecting a particular network for a user based on real-time changes in loads and application requirements. Work is occurring in Release 12 to define parameters that would control switching from LTE to Wi-Fi or from Wi-Fi to LTE.[196]

Bidirectional offloading, however, creates various challenges, as shown in Figure 97 and discussed below.

---

[195] Courtesy Smith Micro Software, 2014. http://www.smithmicro.com.

[196] 3GPP, *Study on Wireless Local Area Network (WLAN) - 3GPP radio interworking (Release 12)*, TR 37.834.
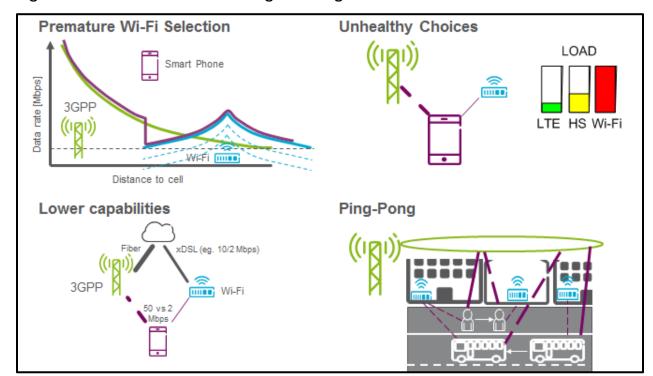
Figure 97: Bidirectional Offloading Challenges

- ❑ **Premature Wi-Fi Selection.** As Wi-Fi-capable devices move into Wi-Fi coverage, they can prematurely reselect to Wi-Fi without comparative evaluation of existing cellular and incoming Wi-Fi capabilities, possibly resulting in the degradation of the end user experience. Real-time throughput-based traffic steering can mitigate this effect.

- ❑ **Unhealthy choices.** In a mixed network of LTE, HSPA, and Wi-Fi, reselection can occur due to a strong Wi-Fi network signal even though the network is under heavy load. The resulting "unhealthy" choice degrades the end user experience because the performance on the cell edge of a lightly loaded cellular network may be superior to that of the heavily loaded Wi-Fi network. Real-time load-based traffic steering can be beneficial in this scenario.

- ❑ **Lower capabilities.** In some cases, selection to a Wi-Fi network may result in reduced performance even if it offers a strong signal because of other factors, such as lower-bandwidth backhaul. Evaluation of criteria beyond wireless capabilities prior to access selection can improve this circumstance.

- ❑ **Ping-Pong.** Ping-ponging between Wi-Fi and cellular, especially if both offer similar signal strengths, can also degrade the user experience. Hysteresis approaches, similar to those used in cellular inter-radio transfer, can better manage transfer between Wi-Fi and cellular accesses.

3GPP RAN2 is discussing real-time or near-real-time methods to address the challenges discussed above.

Other Integration Technologies (SIPTO, LIPA, IFOM, MAPCON)

Release 10 defines additional options for Wi-Fi integration, including Selected IP Traffic Offload (SIPTO), Local IP Access (LIPA), Multi-Access PDN Connectivity (MAPCON), and IP Flow and Seamless Offload (IFOM).

SIPTO is mostly a mechanism to offload traffic that does not need to flow through the core, such as internet-destined traffic. SIPTO can operate on a home femtocell, or it can operate in the macro network.

Local IP Access (LIPA) provides access to local networks, useful with femtocells that normally route all traffic back to the operator network. With LIPA, the UE in a home environment can access local printers, scanners, file servers, media servers, and other resources.

IFOM, as shown in Figure 98, enables simultaneous cellular and Wi-Fi connections, with different traffic flowing over the different connections. A Netflix movie could stream over Wi-Fi, while a VoIP call might flow over the cellular-data connection. IFOM requires the UE to implement Dual Stack Mobile IPv6 (DSMIPv6).

Figure 98: 3GPP IP Flow and Seamless Mobility
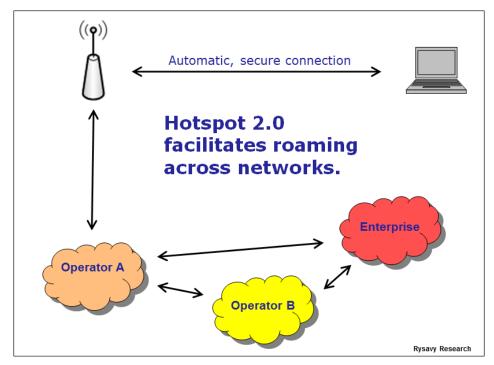
Similar to IFOM, Release 10 feature MAPCON allows multiple simultaneous PDN connections (each with a separate APN), such as Wi-Fi and 3GPP radio access. The UE uses separate IP addresses for each connection but does not need Dual Stack Mobile IPv6 (DSMIPv6).

Hotspot 2.0

Developed by the Wi-Fi Alliance, Hotspot 2.0 specifications, also called "Next Generation Hotspot," facilitate Wi-Fi roaming. Using the IEEE 802.11u standard that allows devices to determine what services are available from an access point, Hotspot 2.0 simplifies the process by which users connect to hotspots, automatically identifying roaming partnerships

and simplifying authentication and connections, as shown in Figure 99.[197] It also provides for encrypted communications over the radio link.[198]

Figure 99: Roaming Using Hotspot 2.0
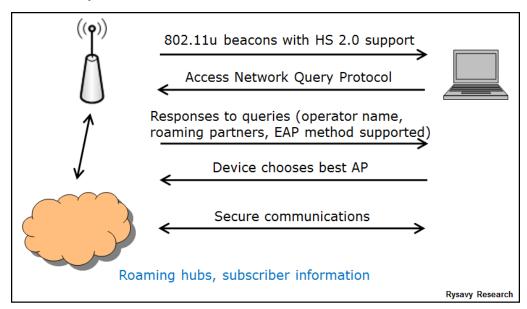
Using IEEE 802.11u, devices can determine what roaming relationships an access point supports and can then securely connect to the Wi-Fi network using one of these roaming arrangements, as shown in Figure 100. Hotspot 2.0 authentication is based on the Extended Authentication Protocol (EAP) using SIM credentials. There are plans to enhance the Hotspot 2.0 protocols in Phase 2, which will define online signup to enable non-SIM-based devices to easily and securely register for services. The Wi-Fi Alliance began a Hotspot 2.0 certification process for devices and access points in June 2012 and uses the designation "Wi-Fi Certified Passpoint" for compliant devices.

---

[197] For example, user devices can be authenticated based on their SIM credentials. Or, users can register or click through an agreement and then not need to redo that with future associations.

[198] The IEEE 802.11i standard has provided encryption for 802.11 communications for many years; however, most hotspots have not implemented this encryption, whereas Hotspot 2.0 does.

Figure 100: Hotspot 2.0 Connection Procedure

Release 2 of Passpoint, available in 2014, added immediate account provisioning, which facilitates a user establishing an account at the point of access. The new version also provides for policies to be downloaded from the network operator; these policies control network selection priorities when multiple networks are available.

Self-Organizing Networks (SON)

As the number of base stations increase through denser deployments and through deployment of femtocells and picocells, manual configuration and maintenance of this infrastructure becomes impractical. With SON, base stations organize and configure themselves by communicating with one another and with the core network. SONs can also self-heal in failure situations.

3GPP began standardization of self-optimization and self-organization in Releases 8 and 9, a key goal being support of multi-vendor environments. Successive releases have augmented SON capabilities.

Features being defined in SON include:

- ❑ Automatic inventory;
- ❑ Automatic software download;
- ❑ Automatic neighbor relation;
- ❑ Automatic physical Cell ID assignment;
- ❑ Mobility robustness/handover optimization;
- ❑ Random access channel optimization;
- ❑ Load-balancing optimization;
- ❑ Inter-cell interference coordination (ICIC) management;

- ❑ Enhanced inter-cell interference coordination (eICIC) management;

- ❑ Coverage and capacity optimization;

- ❑ Cell outage detection and compensation;

- ❑ Self-healing functions;

- ❑ Minimization of drive testing;

- ❑ Energy savings; and

- ❑ Coordination among various SON functions.

3GPP categorizes SON as centralized, distributed, or hybrid, which is a combination of centralized and distributed approaches.

In a centralized architecture, SON algorithms operate on a central network management system or central SON server. In contrast, in a distributed approach, the SON algorithms operate at the eNBs, which make autonomous decisions based on local measurements as well as from other nearby eNBs received via an X2 interface that interconnects eNBs.

The distributed architecture permits faster and easier deployment but is not necessarily as efficient or as consistent in operation, especially in multi-vendor infrastructure deployments.

In a hybrid approach, shown in Figure 101, SON algorithms operate both at the eNB and at a central SON server, with the server supplying values of initial parameters, for example. The eNBs may then update and refine those parameters in response to local measurements.

The hybrid approach resolves deployment scenarios that cannot be resolved by dSON, for example, cases such as:

- ❑ No X2 interface between the eNBs.

- ❑ Multi-vendor deployment with different dSON algorithms.

- ❑ Multi-technology load balancing and user steering.
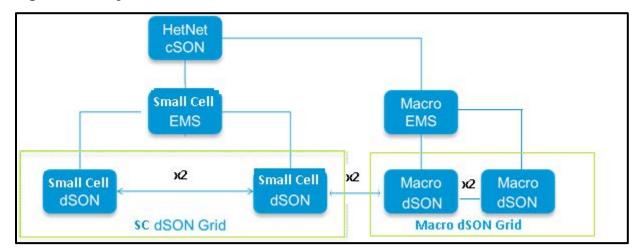
Figure 101: Hybrid SON Architecture[199]

With increasing numbers of macro cells and small cells, interference opportunities increase as well. Optimizing power settings through intelligent power management algorithms is crucial for maximum efficiency with the least amount of interference, including pilot pollution. Pilot pollution can result in low data rates and ping-pong handovers due to channel fading. A hybrid SON approach is well suited for optimized power management.

IP Multimedia Subsystem (IMS)

IP Multimedia Subsystem (IMS) is a service platform for IP multimedia applications: video sharing, PoC, VoIP, streaming video, interactive gaming, and others. IMS by itself does not provide all these applications. Rather, it provides a framework of application servers, subscriber databases, and gateways to make them possible. The exact services will depend on cellular operators and the application developers that make these applications available to operators. The primary application in 2016, however, is VoLTE.

The core networking protocol used within IMS is Session Initiation Protocol (SIP), which includes the companion Session Description Protocol (SDP) used to convey configuration information such as supported voice codecs. Other protocols include Real Time Transport Protocol (RTP) and Real Time Streaming Protocol (RTSP) for transporting actual sessions. The QoS mechanisms in UMTS will be an important component of some IMS applications.

Although originally specified by 3GPP, numerous other organizations around the world are supporting IMS. These include the IETF, which specifies key protocols such as SIP, and the Open Mobile Alliance, which specifies end-to-end, service-layer applications. Other organizations supporting IMS include the GSMA, ETSI, CableLabs, 3GPP2, The Parlay Group, the ITU, ANSI, the Telecoms and Internet Converged Services and Protocols for Advanced Networks (TISPAN), and the Java Community Process (JCP).
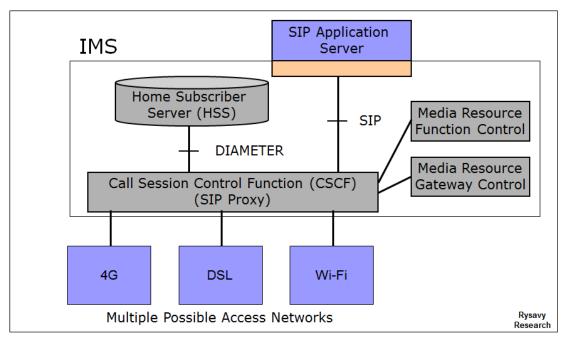
IMS is relatively independent of the radio-access network and can, and likely will, be used by other radio-access networks or wireline networks. Other applications include picture and video sharing that occur in parallel with voice communications. Operators looking to roll out VoIP over networks will use IMS. For example, VoLTE depends on IMS infrastructure.

---

[199] 5G Americas member contribution.

3GPP initially introduced IMS in Release 5 and has enhanced it in each subsequent specification release.

As shown in Figure 102, IMS operates just outside the packet core.

Figure 102: IP Multimedia Subsystem

The benefits of using IMS include handling all communication in the packet domain, tighter integration with the internet, and a lower cost infrastructure based on IP building blocks for both voice and data services.

IMS applications can reside either in the operator's network or in third-party networks including those of enterprises. By managing services and applications centrally—and independently of the access network—IMS can enable network convergence. This allows operators to offer common services across 3G, Wi-Fi, and wireline networks.

Service Continuity, defined in Release 8, provided for a user's entire session to continue seamlessly as the user moves from one access network to another. Release 9 expanded this concept to allow sessions to move across different device types. For example, the user could transfer a video call in midsession from a mobile phone to a large-screen TV, assuming both have an IMS appearance in the network.

Release 8 introduced the IMS Centralized Services (ICS) feature, which allows for IMS-controlled voice features to use either packet-switched or circuit-switched access.

Given that LTE operators will integrate their 5G networks with their current LTE networks, operators are likely to keep using IMS in conjunction with LTE for their voice and other services that use IMS, even as they begin deploying 5G.

Broadcast/Multicast Services

An important capability for 3G and evolved 3G systems is broadcasting and multicasting, wherein multiple users receive the same information using the same radio resource. This

creates a more efficient approach to deliver video when multiple users desire the same content simultaneously. In a broadcast, every subscriber unit in a service area receives the information, whereas in a multicast, only users with subscriptions receive the information. Service areas for both broadcast and multicast can span either the entire network or a specific geographical area. Potential applications include sporting events, select news, venue-specific (shopping mall, museum) information, and even delivery of software upgrades. Giving users the ability to store and replay select content could further expand the scope of applications.

3GPP defined highly efficient broadcast/multicast capabilities for UMTS in Release 6 with MBMS. Release 7 defined optimizations through a feature called multicast/broadcast, single-frequency network operation that involves simultaneous transmission of the exact waveform across multiple cells. This enables the receiver to constructively superpose multiple MBMS Single Frequency Network (SFN), or MBSFN, cell transmissions. The result is highly efficient, WCDMA-based broadcast transmission technology that matches the benefits of OFDMA-based broadcast approaches.

LTE also has a broadcast/multicast capability called eMBMS. OFDM is particularly well suited for efficient broadcasting, as shown in Figure 103, because the mobile system can combine the signal from multiple base stations, also an MBSFN approach, and because of the narrowband nature of OFDM. Normally, these signals would interfere with one another. The single frequency network is a cluster of cells that transmit the same content synchronously with a common carrier frequency.
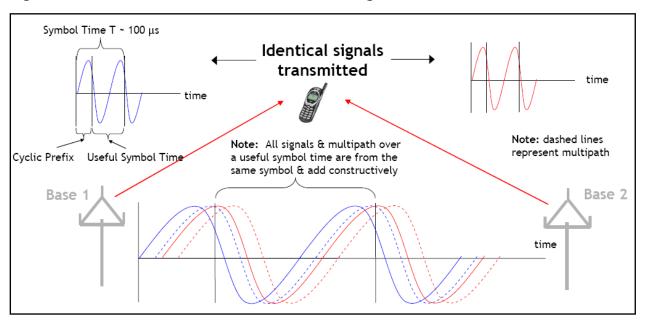
Figure 103: OFDM Enables Efficient Broadcasting

Despite various broadcast technologies being available, market adoption to date has been relatively slow. Internet trends have favored unicast approaches, with users viewing videos of their selection on demand, but there is increasing interest in using eMBMS with LTE to alleviate capacity demands.

Backhaul

Connecting sites to core networks remains a challenge, whether for small cells or macro cells, especially as networks need to deliver higher bandwidth. Fiber is the gold standard, but it is not available everywhere and can be expensive, so operators use a combination of wired and wireless links.

Today's backhaul requirements for LTE can range from 1 to 10 Gbps. By 2020, backhaul requirements could exceed 10 Gbps.[200]

Table 29 and Table 30 summarize the methods and capabilities of the various available approaches.

Table 29: Wired Backhaul Methods and Capabilities[201]

Technology	Distance	Throughput Speed
Direct Fiber	80 km	Hundreds of Mbps to Gbps
Bonded VDSL2	To 5,000 feet	75 Mbps down, 12 Mbps up
FTTX	Most urban areas	Up to 2.5 Gbps down, 1.5 Gbps up
DOCSIS	Most urban areas	Up to 285 Mbps down, 105 Mbps up

Table 30: Wireless Backhaul Methods and Capabilities[202]

Technology	Distance	Line-of-Sight	Throughput Speed
5G Integrated Access and Backhaul	1 km	Yes	1 to 10 Gbps
Millimeter Wave (60 GHz)	1 km	Yes	1 Gbps
Millimeter Wave (70-80 GHz)	3 km (with speed tradeoff)	Yes	10 Gbps
Microwave (6-60 GHz)	Varies by frequency: 2-4 km typical at 30-42 GHz	Yes	1 Gbps+

---

[200] Arthur D. Little, *Creating a Gigabit Society – The Rule of 5G; A report by Arthur D. Little for Vodafone Group,* 2017. See Figure 6.

[201] Small Cell Forum, "Backhaul Technologies for Small Cells," February 2013.

[202] Ibid.

Technology	Distance	Line-of-Sight	Throughput Speed
Licensed sub 6 GHz	1.5 to 10 km	No	170 Mbps (20 MHz TDD), 400 Mbps+ with new technology
Unlicensed sub-6 GHz	Up to 250 meters	No	450 Mbps (IEEE 802.11n 3X3 MIMO)
TV White Space (802.11af-based)	1 to 5 km max throughput, 10 km+ possible	Depends on deployment model	80 Mbps in 6 MHz TDD with 4X4 MIMO
Satellite	Available everywhere	Yes	Up to 50 Mbps downlink, 15 Mbps uplink

UMTS-HSPA

UMTS technology is mature and benefits from research and development that began in the early 1990s. It has been thoroughly trialed, tested, and commercially deployed. UMTS employs a wideband CDMA radio-access technology. The primary benefits of UMTS include high spectral efficiency for voice and data, simultaneous voice and data capability, high user densities that can be supported with low infrastructure costs, and support for high-bandwidth data applications. Operators can also use their entire available spectrum for both voice and high-speed data services.

Additionally, operators can use a common core network, called the UMTS multi-radio network as shown in Figure 104, which supports multiple radio-access networks including GSM, EDGE, WCDMA, HSPA, and evolutions of these technologies.

Figure 104: UMTS Multi-radio Network

HSPA refers to networks that support both HSDPA and HSUPA. All new deployments today are HSPA, and many operators have upgraded their HSDPA networks to HSPA. For example, in 2008, AT&T upgraded most of its network to HSPA. By the end of 2008, HSPA was deployed throughout the Americas.

The UMTS radio-access network consists of base stations referred to as Node B (corresponding to GSM base transceiver systems) that connect to RNCs (corresponding to GSM base station controllers [BSCs]). The RNCs connect to the core network as do the BSCs. When both GSM and WCDMA access networks are available, the network can hand users over between these networks. This is important for managing capacity, as well as in areas in which the operator has continuous GSM coverage, but has only deployed WCDMA in some locations.

Whereas GSM can effectively operate like a spread-spectrum system[203], based on time division in combination with frequency hopping, WCDMA is a direct-sequence, spread-spectrum system. WCDMA is spectrally more efficient than GSM, but it is the wideband nature of WCDMA that provides its greatest advantage—the ability to translate the available spectrum into high data rates. This wideband technology approach results in the flexibility to manage multiple traffic types including voice, narrowband data, and wideband data.

HSDPA

HSDPA, specified in 3GPP Release 5, saw the introduction of high-performance, packet data service that delivers peak theoretical rates of 14 Mbps. Peak user-achievable throughput rates in initial deployments are well over 1 Mbps and as high as 4 Mbps in some networks. The same radio carrier can simultaneously service UMTS voice and data users, as well as HSDPA data users.

HSDPA achieves its high speeds through techniques similar to those that push EDGE performance past GPRS including higher order modulation, variable coding, and soft combining, as well as through the addition of fast scheduling and other techniques.

HSDPA achieves its performance gains from the following radio features:

- ❑ High-speed channels shared in both code and time domains
- ❑ Short TTI
- ❑ Fast scheduling and user diversity
- ❑ Higher order modulation
- ❑ Fast link adaptation
- ❑ Fast HARQ

These features function as follows:

High-Speed Shared Channels and Short Transmission Time Interval: First, HSDPA uses high-speed data channels called "High Speed Physical Downlink Shared Channels" (HS-PDSCH). Up to 15 of these channels can operate in the 5 MHz WCDMA radio channel. Each uses a fixed spreading factor of 16. User transmissions are assigned to one or more of these channels for a short TTI of 2 msec. The network can then readjust how users are assigned to different HS-PDSCH every 2 msec. Resources are thus assigned in both time

---

[203] Spread spectrum systems can either be direct sequence or frequency hopping.

(the TTI interval) and code domains (the HS-PDSCH channels). Figure 105 illustrates different users obtaining different radio resources.
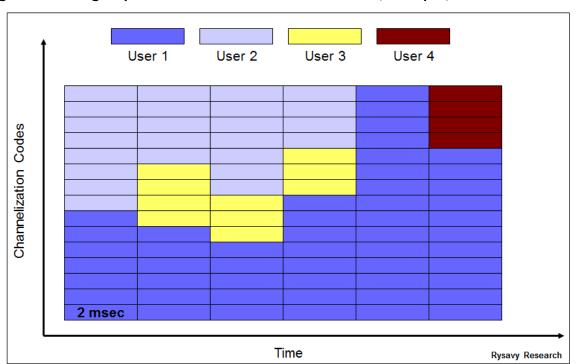
Figure 105: High Speed–Downlink Shared Channels (Example)

Rysavy Research

Fast Scheduling and User Diversity: Fast scheduling exploits the short TTI by assigning users channels that have the best instantaneous channel conditions, rather than in a round-robin fashion. Because channel conditions vary somewhat randomly across users, most users can be serviced with optimum radio conditions and thereby obtain optimum data throughput. Figure 106 shows how a scheduler might choose between two users based on their varying radio conditions to emphasize the user with better instantaneous signal quality. With about 30 users active in a sector, the network achieves significant user diversity and much higher spectral efficiency. The system also ensures that each user receives a minimum level of throughput, an approach called proportional fair scheduling.
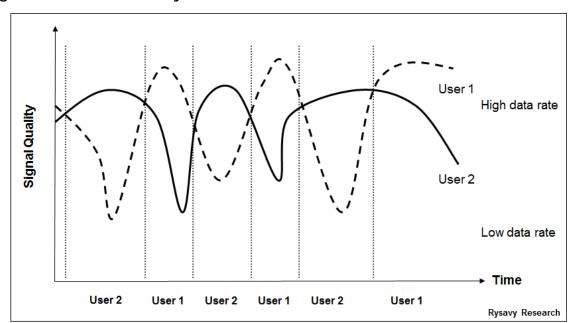
Figure 106: User Diversity

Higher Order Modulation: HSDPA uses both the modulation used in WCDMA—namely QPSK—and, under good radio conditions, an advanced modulation scheme—16 QAM. 16 QAM transmits 4 bits of data in each radio symbol compared to 2 bits with QPSK. Data throughput is increased with 16 QAM, while QPSK is available for adverse radio conditions. HSPA Evolution adds 64 QAM modulation to further increase throughput rates. 64 QAM became available in Release 7, and the combination of MIMO and 64 QAM became available in Release 8.

Fast Link Adaptation: Depending on the condition of the radio channel, different levels of forward-error correction (channel coding) can also be employed. For example, a three-quarter coding rate means that three quarters of the bits transmitted are user bits, and one quarter are error-correcting bits. Fast link adaptation refers to the process of selecting and quickly updating the optimum modulation and coding rate and occurs in coordination with fast scheduling.

Fast Hybrid Automatic Repeat Request: Another HSDPA technique is Fast Hybrid Automatic Repeat Request (Fast Hybrid ARQ). "Fast" refers to the medium-access control mechanisms implemented in Node B (along with scheduling and link adaptation), as opposed to the BSC in GPRS/EDGE, and "hybrid" refers to a process of combining repeated data transmissions with prior transmissions to increase the likelihood of successful decoding. Managing and responding to real-time radio variations at the base station, as opposed to an internal network node, reduces delays and further improves overall data throughput.

Using the approaches just described, HSDPA maximizes data throughputs and capacity and minimizes delays. For users, this translates to better network performance under loaded conditions, faster application performance, and a greater range of applications that function well.

Field results validate the theoretical throughput results. With initial 1.8 Mbps peak rate devices, vendors measured consistent throughput rates in actual deployments of more than 1 Mbps. These rates rose to more than 2 Mbps for 3.6 Mbps devices and then close to 4 Mbps for 7.2 Mbps devices.

In 2008, typical devices supporting peak data rates of 3.6 Mbps or 7.2 Mbps became available. Many operator networks support 7.2 Mbps peak operation, and some even support the maximum rate of 14.4 Mbps.

HSUPA

Whereas HSDPA optimizes downlink performance, HSUPA—which uses the Enhanced Dedicated Channel (E-DCH)—constitutes a set of improvements that optimizes uplink performance. Networks and devices supporting HSUPA became available in 2007. These improvements include higher throughputs, reduced latency, and increased spectral efficiency. HSUPA was standardized in Release 6. It results in an approximately 85% increase in overall cell throughput on the uplink and more than a 50% gain in user throughput. HSUPA also reduces packet delays, a significant benefit resulting in much improved application performance on HSPA networks

Although the primary downlink traffic channel supporting HSDPA serves as a shared channel designed for the support of services delivered through the packet-switched domain, the primary uplink traffic channel defined for HSUPA is a dedicated channel that could be used for services delivered through either the circuit-switched or the packet-switched domains. Nevertheless, by extension and for simplicity, the WCDMA-enhanced uplink capabilities are often identified in the literature as HSUPA.

HSUPA achieves its performance gains through the following approaches:

- ❑ An enhanced dedicated physical channel.

- ❑ A short TTI, as low as 2 msec, which allows faster responses to changing radio conditions and error conditions.

- ❑ Fast Node B-based scheduling, which allows the base station to efficiently allocate radio resources.

- ❑ Fast Hybrid ARQ, which improves the efficiency of error processing.

The combination of TTI, fast scheduling, and Fast Hybrid ARQ also serves to reduce latency. HSUPA can operate with or without HSDPA in the downlink, although use the two approaches together. The improved uplink mechanisms also translate to better coverage and, for rural deployments, larger cell sizes.

HSUPA can achieve different throughput rates based on various parameters including the number of codes used, the spreading factor of the codes, the TTI value, and the transport block size in bytes.

Initial devices enabled peak user rates of close to 2 Mbps as measured in actual network deployments, while current devices have throughputs of more than 5 Mbps. Future devices could have network rates as high as 69 Mbps, as discussed further below.

Beyond throughput enhancements, HSUPA also significantly reduces latency.

Evolution of HSPA (HSPA+)

The goal in evolving HSPA is to exploit available radio technologies—largely enabled by increases in digital signal processing power—to maximize CDMA-based radio performance. This evolution has significantly advanced HSPA and extends the life of sizeable operator infrastructure investments.

Wireless and networking technologists have defined a series of enhancements for HSPA, beginning in Release 7 and now continuing through Release 14. These include advanced receivers, multi-carrier operation, MIMO, Continuous Packet Connectivity, Higher-Order Modulation, One-Tunnel Architecture, HetNet support, and advanced voice capabilities both in circuit- and packet-switched domains.

Taking advantage of these various radio technologies, 3GPP has standardized a number of features, beginning in Release 7 including higher order modulation and MIMO. Collectively, these capabilities are referred to as HSPA+. Release 8 through Release 12 include further enhancements.

The goals of HSPA+ were to:

❑ Exploit the full potential of a CDMA approach.

❑ Provide smooth interworking between HSPA+ and LTE, thereby facilitating the operation of both technologies. As such, operators may choose to leverage the EPC planned for LTE.

❑ Allow operation in a packet-only mode for both voice and data.

❑ Be backward-compatible with previous systems while incurring no performance degradation with either earlier or newer devices.

❑ Facilitate migration from current HSPA infrastructure to HSPA+ infrastructure.

HSPA improvements have continued through successive 3GPP releases, including Release 14, which has downlink interference mitigation. Release 15 has work items for quality of experience, multi-carrier enhancements, and various protocol enhancements.

The following sections discuss specific enhancements that have already been implemented in HSPA.

Advanced Receivers

3GPP has specified a number advanced-received designs including: Type 1, which uses mobile-receive diversity; Type 2, which uses channel equalization; and Type 3, which includes a combination of receive diversity and channel equalization. Type 3i devices, which became available in 2012, employ interference cancellation. Note that the different types of receivers are release-independent. For example, Type 3i receivers will work and provide a capacity gain in an earlier Release 5 network.

The first approach is mobile-receive diversity. This technique relies on the optimal combination of received signals from separate receiving antennas. The antenna spacing yields signals that have somewhat independent fading characteristics. Hence, the combined signal can be more effectively decoded, which almost doubles downlink capacity when done in combination with channel equalization. Receive diversity is effective even with smaller devices such as like PC Card modems and smartphones.

Current receiver architectures based on rake receivers are effective for speeds up to a few megabits per second. But at higher speeds, the combination of reduced symbol period and multipath interference results in Intersymbol Interference and diminishes rake receiver performance. This problem can be solved by advanced-receiver architectures with channel equalizers that yield additional capacity gains over HSDPA with receive diversity. Alternate advanced-receiver approaches include interference cancellation and generalized rake receivers (G-Rake). Different vendors are emphasizing different approaches. The performance requirements for advanced-receiver architectures, however, were specified in 3GPP Release 6. The combination of mobile-receive diversity and channel equalization (Type 3) is especially attractive, because it results in a large capacity gain independent of the radio channel.

What makes such enhancements attractive is that the networks do not require any changes other than increased capacity within the infrastructure to support the higher bandwidth. Moreover, the network can support a combination of devices including both earlier devices that do not include these enhancements and later devices that do. Device vendors can selectively apply these enhancements to their higher-end devices.

MIMO

Another standardized capability is MIMO, a technique that employs multiple transmit antennas and multiple receive antennas, often in combination with multiple radios and multiple parallel data streams. The most common use of the term "MIMO" applies to spatial multiplexing. The transmitter sends different data streams over each antenna. Whereas multipath is an impediment for other radio systems, MIMO—as illustrated in Figure 107—actually exploits multipath, relying on signals to travel across different uncorrelated communications paths. The multiple data paths effectively operate in parallel and, with appropriate decoding, in a multiplicative gain in throughput.
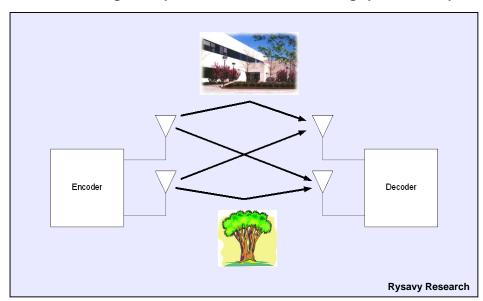
Figure 107: MIMO Using Multiple Paths to Boost Throughput and Capacity

Tests of MIMO have proven effective in WLANs operating in relative isolation where interference is not a dominant factor. Spatial multiplexing MIMO can also benefit HSPA "hotspots" serving local areas including airports, campuses, and malls. In a fully loaded network with interference from adjacent cells, however, overall capacity gains will be more

modest—in the range of 20% to 33% over mobile-receive diversity. Relative to a 1x1 antenna system, however, 2X2 MIMO can deliver cell throughput gains of about 80%. 3GPP has standardized spatial multiplexing MIMO in Release 7 using Double Transmit Adaptive Array (D-TxAA).

Release 9 provides for a means to leverage MIMO antennas at the base station when transmitting to user equipment that does not support MIMO. The two transmit antennas in the base station can transmit a single stream using beam forming. This is called "single stream MIMO" or "MIMO with single-stream restriction" and results in higher throughput rates because of the improved signal received by the user equipment.

3GPP designed uplink dual-antenna beamforming and 2X2 MIMO for HSPA+ in Release 11.

Continuous Packet Connectivity

Continuous Packet Connectivity (CPC) specified in Release 7 reduces the uplink interference created by the dedicated physical control channels of packet data users when those channels have no user data to transmit, which increases the number of simultaneously connected HSUPA users. CPC allows both discontinuous uplink transmission and discontinuous downlink reception, wherein the modem can turn off its receiver after a certain period of HSDPA inactivity. CPC is especially beneficial to VoIP on the uplink because the radio can turn off between VoIP packets, as shown in Figure 108.
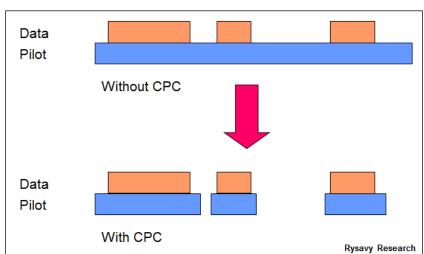
Figure 108: Continuous Packet Connectivity

Higher Order Modulation

Another way of increasing performance is with higher order modulation. HSPA uses 16 QAM on the downlink and QPSK on the uplink, but HSPA+ adds 64 QAM to the downlink and 16 QAM to the uplink. 3GPP has also introduced 64 QAM to the uplink for HSPA+ in Release 11. Higher order modulation requires a better SNR, achieved through receive diversity and equalization.

Multi-Carrier HSPA

3GPP defined dual-carrier HSPA operation in Release 8, which coordinates the operation of HSPA on two adjacent 5 MHz carriers so that data transmissions can achieve higher

throughput rates, as shown in Figure 109. The work item assumed two adjacent carriers, downlink operation and no MIMO. This configuration achieves a doubling of the 21 Mbps maximum rate available on each channel to 42 Mbps.
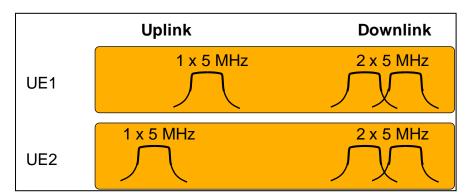
Figure 109: Dual-Carrier Operation with One Uplink Carrier[204]

Benefits include:

❑ An increase in spectral efficiency of about 15%, comparable to what can be obtained with 2X2 MIMO.

❑ Significantly higher peak throughputs available to users, especially in lightly-loaded networks.

❑ Same maximum-throughput rate of 42 Mbps as using MIMO, but with a less expensive infrastructure upgrade.

Scheduling packets across two carriers is a more efficient use of resources, resulting in what is called "trunking gain." Multi-user diversity also improves from an increased number of users across the two channels.

Release 9 also supports dual-carrier operation in the uplink. Release 10 specifies the use of up to four channels, resulting in peak downlink data rates of 168 Mbps. Release 11 supports eight radio channels on the downlink, resulting in a further doubling of theoretical throughput to 336 Mbps. On the uplink, devices can transmit using two antennas for either rank 1 (single stream beamforming) or rank 2 (dual-stream MIMO) transmission modes. Rank 1 beamforming helps with coverage (approximately 40%), while rank 2 MIMO helps with throughput speeds (approximately 20% median and 80% at cell edge). In addition, 64 QAM will be possible on the uplink, enabling uplink speeds to 69 Mbps in dual-carrier operation.

Downlink Multiflow Transmission

Release 11 specifies means by which two cells can transmit to the mobile station at the same time. The two cells transmit independent data, in effect a spatial multiplexing approach, improving both peak and average data.

---

[204] Harri Holma and Antti Toskala, *LTE for UMTS, OFDMA and SC-FDMA Based Radio Access*, Wiley, April 2009.

Multiflow transmission with HSPA+ also enhances HetNet operation in which picocell coverage can be expanded within a macrocell coverage area, as shown in Figure 110.

Figure 110: HSPA+ HetNet Using Multipoint Transmission[205]

Multiflow enhances HSPA+ network operation using the following approaches:

- ❑ **Single Frequency Dual Cell.** The UE communicates with two different cells using the same frequency, improving cell-edge performance and providing network load balancing.

- ❑ **Dual Frequency Three Cell.** The UE communicates with two different cells using the same frequency. In addition, it communicates with one other cell on a different frequency.

- ❑ **Dual Frequency Four Cells.** The UE communicates using two instances of Single Frequency Dual Cell operation as described above.

In Release 12, 3GPP is considering the following enhancement to Multiflow operation, which is primarily targeted towards HetNet operation:

- ❑ **Dual Frequency Dual Carrier.** The UE aggregates cells on two different frequencies from two different sites.

HSPA+ Throughput Rates

Table 31 summarizes the capabilities of HSPA and HSPA+ based on the various methods discussed above.

---

[205] Qualcomm, "HSPA+ Advanced: Taking HSPA+ to the Next Level," February 2012, http://www.qualcomm.com/media/documents/hspa-advanced-taking-hspa-next-level-whitepaper, accessed June 20, 2014.

Table 31: HSPA Throughput Evolution

Technology	Downlink (Mbps) Peak Data Rate	Uplink (Mbps) Peak Data Rate
HSPA as defined in Release 6	14.4	5.76
Release 7 HSPA+ DL 64 QAM, UL 16 QAM, 5+5 MHz	21.1	11.5
Release 7 HSPA+ 2X2 MIMO, DL 16 QAM, UL 16 QAM, 5+5 MHz	28.0	11.5
Release 8 HSPA+ 2X2 MIMO DL 64 QAM, UL 16 QAM, 5+5 MHz	42.2	11.5
Release 8 HSPA+ (no MIMO) Dual Carrier, 10+5 MHz	42.2	11.5
Release 9 HSPA+ 2X2 MIMO, Dual Carrier DL and UL, 10+10 MHz	84.0	23.0
Release 10 HSPA+ 2X2 MIMO, Quad Carrier[206] DL, Dual Carrier UL, 20+10 MHz	168.0	23.0
Release 11 HSPA+ 2X2 MIMO DL and UL, 8 Carrier DL, Dual Carrier UL, 40+10 MHz	336.0	69.0

Release 13 enables aggregation of two UL carriers across bands.

Figure 111 shows the cumulative distribution function of throughput values in a commercially deployed Release 8 HSPA+ network in an indoor coverage scenario. The figure shows significant performance gains from higher-order modulation and MIMO.

---

[206] No operators have announced plans to deploy HSPA in a quad (or greater) carrier configuration. Three carrier configurations, however, have been deployed.

Figure 111: HSPA+ Performance Measurements Commercial Network (5+5 MHz)[207]

The figure shows a reasonably typical indoor scenario in a macro-cell deployment. Under better radio conditions, HSPA+ will achieve higher performance results.

Figure 112 shows the benefit of dual-carrier operation (no MIMO employed), which essentially doubles throughputs over single carrier operation.

---

[207] 5G Americas member company contribution.

Figure 112: Dual-Carrier HSPA+ Throughputs[208]

HSPA+ also has improved latency performance of as low as 25 msec and improved packet call setup time of below 500 msec.

Figure 113 summarizes the key capabilities and benefits of the features being deployed in HSPA+.

---

[208] 5G Americas member company contribution. 64 QAM.

Figure 113: Summary of HSPA Functions and Benefits[209]

Uplink DTX + downlink DRX	→ Lower UE power consumption
CS voice over HSPA, VoIP, WCDMA+	→ Higher voice capacity or more capacity for data
Downlink 64QAM, MIMO, and multi carrier	→ Higher downlink peak data rates and higher data capacity
Uplink 16QAM, MIMO, and dual carrier	→ Higher uplink peak data rates and higher data capacity
Het-net support, pico range expansion	→ Higher network capacity
High speed FACH, High speed RACH, FE-FACH	→ Lower latency = better response times / More efficient common channels = savings in channel elements
Flat architecture optimization	→ Fewer network elements

UMTS TDD and TD-SCDMA

Most WCDMA and HSDPA deployments are based on FDD, which uses different radio bands for transmit and receive. In the alternate TDD approach, transmit and receive functions alternate in time on the same radio channel. 3GPP specifications include a TDD version of UMTS, called "UMTS TDD."

TDD does not provide any inherent advantage for voice functions, which need balanced links—namely, the same amount of capacity in both the uplink and the downlink. Many data applications, however, are asymmetric, often with the downlink consuming more bandwidth than the uplink. A TDD radio interface can dynamically adjust the downlink-to-uplink ratio accordingly, hence balancing both forward-link and reverse-link capacity. Note that for UMTS FDD, the higher spectral efficiency achievable in the downlink versus the uplink addresses the asymmetrical nature of average data traffic.

---

[209] 5G Americas member contribution.

The UMTS TDD specification also includes the capability to use joint detection in receiver-signal processing, which offers improved performance.

One consideration, however, relates to available spectrum. Various countries around the world including those in Europe, Asia, and the Pacific region have licensed spectrum available specifically for TDD systems. TDD is also a good choice for any spectrum that does not provide a duplex gap between forward and reverse links.

In the United States, there is limited spectrum specifically allocated for TDD systems, the major band being BRS at 2.5 GHz used by Clearwire for WiMAX and now LTE TDD.[210] UMTS TDD is not a good choice in FDD bands; it would not be able to operate effectively in both bands, thereby making the overall system efficiency relatively poor.

TDD systems require network synchronization and careful coordination between operators or guardbands, which may be problematic in certain bands.

There has not been widespread deployment of UMTS TDD.

Time Division Synchronous Code Division Multiple Access (TD-SCDMA) is one of the official 3G wireless technologies, mostly for deployment in China. Specified through 3GPP as a variant of the UMTS TDD System and operating with a 1.28 megachips per second (Mcps) chip rate versus 3.84 Mcps for UMTS TDD, TD-SCDMA's primary attribute is that it supports very high subscriber densities, making it a possible alternative for wireless local loops. TD-SCDMA uses the same core network as UMTS, and it is possible for the same core network to support both UMTS and TD-SCDMA radio-access networks.

Although there are no planned deployments in any country other than China, TD-SCDMA could theoretically be deployed anywhere unpaired spectrum is available—such as the bands licensed for UMTS TDD—assuming appropriate resolution of regulatory issues.

EDGE/EGPRS

Today, most GSM networks support EDGE, an enhancement to GPRS, which is the original packet data service for GSM networks.[211] GPRS provides a packet-based IP connectivity solution supporting a wide range of enterprise and consumer applications. GSM networks with EDGE operate as wireless extensions to the internet and give users internet access, as well as access to their organizations from anywhere. Peak EDGE user-achievable[212] throughput rates are up to 200 Kbps. Figure depicts the system architecture.

---

[210] The 1910-1920 MHz band targeted unlicensed TDD systems but has never been used.

[211] GSM technology also provides circuit-switched data services, which are not described in this paper since they are seldom used.

[212] "Peak user-achievable" means users, under favorable conditions of network loading and signal propagation, can achieve this rate as measured by applications such as file transfer. Average rates depend on many factors and will be lower than these rates.
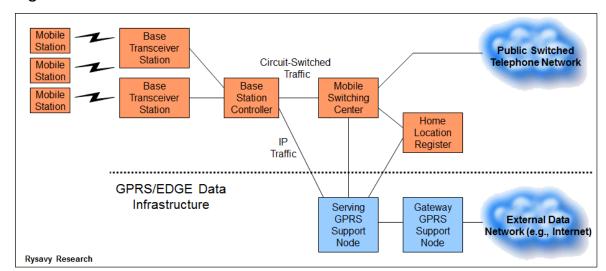
Figure 114: GSM/GPRS/EDGE Architecture

EDGE is essentially the addition of a packet-data infrastructure to GSM. In fact, this same data architecture is preserved in UMTS and HSPA networks, and the data architecture is technically referred to as GPRS for the core-data function in all of these networks. The term GPRS may also be used to refer to the initial radio interface, now supplanted by EDGE. Functions of the data elements are as follows:
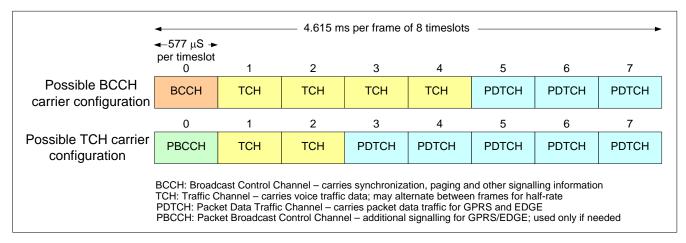
❑ The base station controller directs/receives packet data to/from the Serving GPRS Support Node (SGSN), an element that authenticates and tracks the location of mobile stations.

❑ The SGSN performs the types of functions for data that the Mobile Switching Center (MSC) performs for voice. Each serving area has one SGSN, and it is often collocated with the MSC.

❑ The SGSN forwards/receives user data to/from the Gateway GPRS Support Node (GGSN), which can be viewed as a mobile IP router to external IP networks. Typically, there is one GGSN per external network (for example, the internet). The GGSN also manages IP addresses, dynamically assigning them to mobile stations for their data sessions.

Another important element is the Home Location Register (HLR), which stores users' account information for both voice and data services. Of significance is that this same data architecture supports data services in GSM and in UMTS-HSPA networks, thereby simplifying operator network upgrades.

In the radio link, GSM uses radio channels of 200 kilohertz (kHz) width, divided in time into eight timeslots comprising 577 microseconds (μs) that repeat every 4.6 msec, as shown in Figure . The network can have multiple radio channels (referred to as transceivers) operating in each cell sector. The network assigns different functions to each timeslot such as the Broadcast Control Channel (BCCH), circuit-switched functions like voice calls or data calls, the optional Packet Broadcast Control Channel (PBCCH), and packet data channels. The network can dynamically adjust capacity between voice and data functions, and it can also reserve minimum resources for each service. This scheduling approach enables more data traffic when voice traffic is low or, likewise, more voice traffic when data traffic is low, thereby maximizing overall use of the network. For example, the PBCCH, which expands

the capabilities of the normal BCCH, may be set-up on an additional timeslot of a Time Division Multiple Access (TDMA) frame when justified by the volume of data traffic.

Figure 115: Example of GSM/EDGE Timeslot Structure[213]

EDGE offers close coupling between voice and data services. In most networks, while in a data session, users can accept an incoming voice call, which suspends the data session, and then resume their data session automatically when the voice session ends. Users can also receive SMS messages and data notifications[214] while on a voice call, as described below.

With respect to data performance, each data timeslot can deliver peak user-achievable data rates of up to about 40 Kbps. The network can aggregate up to five timeslots on the downlink and up to four timeslots on the uplink with current devices.

If multiple data users are active in a sector, they share the available data channels. As demand for data services increases, however, an operator can accommodate customers by assigning an increasing number of channels for data service that is limited only by that operator's total available spectrum and radio planning.

EDGE is an official 3G cellular technology that can be deployed within an operator's existing 850, 900, 1800, and 1900 MHz spectrum bands. EDGE capability is now largely standard in new GSM deployments. A GPRS network using the EDGE radio interface is technically called an "Enhanced GPRS" (EGPRS) network, and a GSM network with EDGE capability is referred to as GSM Edge Radio Access Network (GERAN). EDGE has been an inherent part of GSM specifications since Release 99. It is fully backward-compatible with older GSM networks, meaning that GPRS devices work on EDGE networks and that GPRS and EDGE terminals can operate simultaneously on the same traffic channels.

Dual Transfer Mode (DTM) devices can simultaneously communicate voice and data. DTM is a 3GPP-specified technology that enables new applications like video sharing while providing a consistent service experience (service continuity) with UMTS. Typically, a DTM

---

[213] 5G Americas member company contribution.

[214] Example: WAP notification message delivered via SMS.

end-to-end solution requires only a software upgrade to the GSM/EDGE radio network. There are a number of networks and devices supporting DTM.

A feature in Release 9 that applies to EDGE is the Enhanced Flexible Timeslot Assignment (EFTA), which allows for more efficient adaptation to varying uplink versus downlink transmission needs. The network allocates uplink and downlink timeslots that overlap in time, and the mobile station may either use the corresponding uplink timeslots for transmission or receive on the overlapping downlink time slot, if it has nothing to transmit. In addition, alternative EFTA multi-slot classes enable the support of as many as eight timeslots per downlink carrier (instead of five or six timeslots with multi-slot classes 30 to 45).

Abbreviations and Acronyms

The following abbreviations are used in this paper. Abbreviations are defined on first use.

1G – First Generation

1xEV-DO – One Carrier Evolution, Data Optimized

1xEV-DV – One Carrier Evolution, Data Voice

1XRTT – One Carrier Radio Transmission Technology

2G – Second Generation

3G – Third Generation (meeting requirements set forth by the ITU IMT project)

3GPP – Third Generation Partnership Project

3GPP2 – Third Generation Partnership Project 2

4G – Fourth Generation (meeting requirements set forth by the ITU IMT-Advanced project)

5GC – 5G Core

8-PSK – Octagonal Phase Shift Keying

AAS – Adaptive Antenna Systems

ABR – Allocation Retention Priority

AGW – Access Gateway

AMF – Access and Mobility Management Function

AMR – Adaptive Multi Rate

AMR-WB – Adaptive Multi-Rate Wideband

ANDSF – Access Network Discovery and Selection Function.

ANSI – American National Standards Institute

APCO – Association of Public Safety Officials

API – Application Programming Interface

APN – Access Point Name

ARP – Allocation Retention Priority

ARPU – Average Revenue per User

ARQ – Automatic Repeat Request

ATM – Asynchronous Transfer Mode

AWGN – Additive White Gaussian Noise Channel

AWS – Advanced Wireless Services

BCCH – Broadcast Control Channel

bps – bits per second

BRS – Broadband Radio Service

BSC – Base Station Controller

BTS – Base Transceiver Station

C/I – Carrier to Intermodulation Ratio

CAPEX- Capital Expenditure

CBF – Coordinated Beam Forming

CBRS – Citizens Broadband Radio Service

CBS – Coordinated Beam Switching

CSS3 – Cascading Style Sheets 3 (CSS3)

CDD – Cyclic Delay Diversity

CDF – Cumulative Distribution Function

CDMA – Code Division Multiple Access

CL – Closed Loop

CL-SM – Closed Loop Spatial Multiplexing

CMAS – Commercial Mobile Alert System

CMOS – Complementary Metal Oxide Semiconductor

CoMP – Coordinated Multi Point

cMTC – Critical Machine Type Communications

CP – Cyclic Prefix

CPC – Continuous Packet Connectivity

CPRI – Common Public Radio Interface

CQI - Channel Quality Indicators

C-RAN – Cloud Radio Access Network

CRM – Customer Relationship Management

CRS – Cell-specific Reference Signal

CS – Convergence Sublayer

CSFB – Circuit-Switched Fallback

CTIA – Cellular Telephone Industries Association

C-V2X – Cellular Vehicle-to-X

D-AMPS – Digital Advanced Mobile Phone Service

DAS – Distributed Antenna System

DAS – Downlink EGPRS2-A Level Scheme

dB – Decibel

DBS – Downlink EGPRS2-B Level Scheme

DC-HSPA – Dual Carrier HSPA

DFT – Discrete Fourier Transform

DL – Downlink

DNS – Domain Name Service

DPCCH – Dedicated Physical Control Channel

DPS – Dynamic Point Selection

DSL – Digital Subscriber Line

DSMIPv6 – Dual Stack Mobile IPv6

DTM – Dual Transfer Mode

DRX – Discontinuous Reception

D-TxAA – Double Transmit Adaptive Array

DVB-H – Digital Video Broadcasting Handheld

E-DCH – Enhanced Dedicated Channel

EBCMCS – Enhanced Broadcast Multicast Services

EC-GSM – Extended Coverage GSM

eCoMP – enhanced CoMP

EDGE – Enhanced Data Rates for GSM Evolution

EFTA – Enhanced Flexible Timeslot Assignment

EGPRS – Enhanced General Packet Radio Service

eICIC – Enhanced Inter-Cell Interference Coordination

eMBMS – Enhanced Multimedia Broadcast Multicast Services

eNodeB – Evolved Node B

EAP – Extensible Authentication Protocol

eLAA – Enhanced Licensed-Assisted Access

eNB – Evolved Node B

EPC – Evolved Packet Core

EPDCCH – Enhanced Physical Downlink Control Channel

eMBB – Enhanced Mobile Broadband

ePDG – Enhanced Packet Data Gateway

EPS – Evolved Packet System

ERP – Enterprise Resource Planning

eSaMOG – Enhanced S2a-based Mobility over GTP

ESC – Environmental Sensing Capability

eSRVCC – Enhanced Single-Radio Voice Call Continuity

ETRI – Electronic and Telecommunications Research Institute

ETSI – European Telecommunications Standards Institute

E-UTRAN – Enhanced UMTS Terrestrial Radio Access Network

EVS – Enhanced Voice Services (codec)

FE-FACH – Further Enhanced Forward Access Channel

EV-DO –Evolution, Data Optimized

EV-DV –Evolution, Data Voice

EVRC – Enhanced Variable Rate Codec

FBMC – Filter-Bank Multi-Carrier

FCC – Federal Communications Commission

FDD – Frequency Division Duplex

FeCoMP – Further Enhanced Coordinated Multi Point

feICIC – Further enhanced ICIC

FirstNet – First Responder Network Authority

Flash OFDM – Fast Low-Latency Access with Seamless Handoff OFDM

FLO – Forward-Link Only

FMC – Fixed Mobile Convergence

FP7 – Seventh Framework Programme

FTP – File Transfer Protocol

GAA – General Authorized Access

GAN – Generic Access Network

GB – Gigabyte

Gbps – Gigabits Per Second

GBR – Guaranteed Bit Rate

GByte – Gigabyte

GCS – Group Communication Service

GERAN – GSM EDGE Radio Access Network

GFDM – Generalized Frequency Division Multiplexing

GGSN – Gateway GPRS Support Node

GHz — Gigahertz

GMSK – Gaussian Minimum Shift Keying

gNB – NR NodeB

GPRS – General Packet Radio Service

G-Rake – Generalized Rake Receiver

GSM – Global System for Mobile Communications

GSMA – GSM Association

HARQ – Hybrid Automatic Repeat Request

HD – High Definition

HetNet – heterogeneous network

HLR – Home Location Register

Hr – Hour

HSDPA – High Speed Downlink Packet Access

HS-FACH – High Speed Forward Access Channel

HS-PDSCH - High Speed Physical Downlink Shared Channels

HS-RACH – High Speed Reverse Access Channel

HSPA – High Speed Packet Access (HSDPA with HSUPA)

HSPA+ – HSPA Evolution

HSS – Home Subscriber Server

HSUPA – High Speed Uplink Packet Access

Hz – Hertz

ICIC – Inter-Cell Interference Coordination

ICN – Information-Centric Networking

ICS – IMS Centralized Services

ICT – Information and Communication Technologies

IEEE – Institute of Electrical and Electronic Engineers

IETF – Internet Engineering Taskforce

IFFT – Inverse Fast Fourier Transform

IFOM – IP Flow and Seamless Offload

IM – Instant Messaging

IMS – IP Multimedia Subsystem

IMT – International Mobile Telecommunications

IMT-Advanced - International Mobile Telecommunications-Advanced

IRC – Interference Rejection Combining

IoT – Internet of Things

IPR - Intellectual Property Rights

IP – Internet Protocol

IPTV – Internet Protocol Television

IR – Incremental Redundancy

ISD – Inter-site Distance

ISI – Intersymbol Interference

ISP – Internet Service Provider

ITU – International Telecommunication Union

JCP – Java Community Process

JR – Joint Reception

JT – Joint Transmission

Kbps – Kilobits Per Second

kHz — Kilohertz

km – Kilometer

LAA – License-Assisted Access

LBT – Listen-Before-Talk

LDPC - Low-Density Parity Code

LIPA – Local IP Access

LMDS – Local Multipoint Distribution Service

LPWA – Low-Power Wide-Area

LTE – Long Term Evolution

LTE-A – LTE-Advanced

LTE-TDD – LTE Time Division Duplex

LTE-U – LTE-Unlicensed

LSTI – LTE/SAE Trial Initiative

LWA – LTE Wi-Fi Aggregation

LWIP – LTE WLAN Radio Level Integration with IPsec Tunnel

M2M – Machine-to-machine

MAC – Medium-Access Control

MAPCON – Multi-Access PDN Connectivity

MB - Megabyte

MBMS - Multimedia Broadcast/Multicast Service

Mbps – Megabits Per Second

MBR – Maximum Bit Rate

MBSFN – Multicast/broadcast, Single Frequency

MCPA – Mobile Consumer Application Platform

Mcps – Megachips Per Second

MCPTT – Mission-Critical Push-to-Talk

MCS – Modulation and Coding Scheme

MCW – Multiple Codeword

MDT – Minimization of Drive Tests

MEAP – Mobile Enterprise Application Platforms

MEC – Multi-access Edge Computing

MediaFLO – Media Forward Link Only

METIS – Mobile and wireless communications Enablers for the Twenty-twenty Information Society

MHz – Megahertz

MID – Mobile Internet Devices

MIMO – Multiple Input Multiple Output

MMSE – Minimum Mean Square Error

mITF – Japan Mobile IT Forum

MMDS – Multichannel Multipoint Distribution Service

MME – Mobile Management Entity

mMTC – Massive Machine Type Communications

MOS – Mean Opinion Score

MP-QUIC – Multipath Quick UDP Internet Connections

MP-TCP – Multipath TCP

MRxD – Mobile Receive Diversity

ms – millisecond

MS – Mobile Station

MSA – Mobile Service Architecture

MSC – Mobile Switching Center

MTC – Machine Type Communications

MTC-IWF – Machine-Type Communications Interworking Function (MTC-IWF)

msec – millisecond

MU-MIMO – Multi-User MIMO

MUST – Downlink Multiuser Superposition Transmission

NAICS – Network-Assisted Interference Cancellation and Suppression

NB-IoT – Narrowband Internet of Things

NENA – National Emergency Number Association

NGMC – Next Generation Mobile Committee

NGMN – Next Generation Mobile Networks Alliance

NG-RAN – New Generation Radio Access Network

NOMA – Non-Orthogonal Multiple Access

NR – New Radio

NTIA – National Telecommunications and Information Administration

OFDM – Orthogonal Frequency Division Multiplexing

OFDMA – Orthogonal Frequency Division Multiple Access

OL-SM – Open Loop Spatial Multiplexing

OMA – Open Mobile Alliance

ORI – Open Radio Equipment Interface

PA – Priority Access

PAL – Priority Access License

PAR – Peak to Average Ratio

PBCCH – Packet Broadcast Control Channel

PCH – Paging Channel

PCRF – Policy Control and Charging Rules Function

PCS – Personal Communications Service

PDCP – Packet Data Convergence Protocol

PDN – Packet Data Network

PGW – Packet Gateway

PHY – Physical Layer

PMI – Precoding Matrix Indication

PMIPv6 – Proxy Mobile IPv6

PNF – Physical Network Function

PoC – Push-to-Talk Over Cellular

PSH – Packet Switched Handover

PSK – Phase-Shift Keying

QAM – Quadrature Amplitude Modulation

QCI – Quality of Service Class Identifier

QLIC – Quasi-Linear Interference Cancellation

QoS – Quality of Service

QPSK – Quadrature Phase Shift Keying

QUIC – Quick UDP Internet Connections.

RAB – Radio Access Bearer

RAN – Radio Access Network

RCAF – RAN Congestion Awareness Function

RCLWI - RAN Controlled LTE WLAN Interworking

RCS – Rich Communications Suite

REST – Representational State Transfer

RF – Radio Frequency

RLC – Radio Link Control

RNC – Radio Network Controller

ROHC – Robust Header Compression

RRC – Radio Resource Control

RRH – Remote Radio Head

RRU – Remote Radio Unit

RTP – Real Time Transport Protocol

RTSP – Real Time Streaming Protocol

SAE – System Architecture Evolution

SaMOG – S2a-based Mobility over GTP

SAS – Spectrum Access System

SC-FDMA – Single Carrier Frequency Division Multiple Access

SCMA – Sparse Coded Multiple Access

SCRI – Signaling Connection Release Indication

SCW – Single Codeword

SDAP – Service Data Adaptation Protocol

SDMA – Space Division Multiple Access

SDN – Software-Defined Networking

SDP – Session Description Protocol

sec – Second

SFBA – Space Frequency Block Code

SFN – Single Frequency Network

SGSN – Serving GPRS Support Node

SGW – Serving Gateway

SIC – Successive Interference Cancellation

SIM – Subscriber Identity Module

SIMO – Single Input Multiple Output

SINR – Signal to Interference Plus Noise Ratio

SIP – Session Initiation Protocol

SIPTO – Selected IP Traffic Offload

SISO – Single Input Single Output

SMS – Short Message Service

SNR – Signal to Noise Ratio

SON – Self-Organizing Network

SPS – Semi-Persistent Scheduling

SRVCC – Single-Radio Voice Call Continuity

SU-MIMO – Single User MIMO

SVDO – Simultaneous 1XRTT Voice and EV-DO Data

SVLTE – Simultaneous Voice and LTE

TCH – Traffic Channel

TCP/IP – Transmission Control Protocol/IP

TD – Transmit Diversity

TDD – Time Division Duplex

TDMA – Time Division Multiple Access

TD-SCDMA – Time Division Synchronous Code Division Multiple Access

TD-CDMA – Time Division Code Division Multiple Access

TETRA – Terrestrial Trunked Radio

TIA/EIA – Telecommunications Industry Association/Electronics Industry Association

TISPAN – Telecoms and Internet Converged Services and Protocols for Advanced Networks

TSG-RAN – Technical Services Group Radio Access Network

TTI – Transmission Time Interval

UAS – Uplink EGPRS2-A Level Scheme

UBS – Uplink EGPRS2-B Level Scheme

UE – User Equipment

UFMC – Universal Filtered Multi-Carrier

UICC – Universal Integrated Circuit Card

UL – Uplink

UMA – Unlicensed Mobile Access

UMB – Ultra Mobile Broadband

UMTS – Universal Mobile Telecommunications System

UPCON – User-Plane Congestion Management

UPF – User Plane Function

URA-PCH – UTRAN Registration Area Paging Channel

URI – Uniform Resource Identifier

URLLC – Ultra-Reliable and Low Latency Communications

us – Microsecond

USIM – UICC SIM

UTRAN – UMTS Terrestrial Radio Access Network

V2X – vehicle -to-infrastructure

VAMOS – Voice Services over Adaptive Multi-User Channels on One Slot

VDSL – Very-High-Bit-Rate DSL

VEPC – Virtualized EPC

ViLTE – Video Over LTE

VoIP – Voice over Internet Protocol

VoHSPA – Voice over HSPA

VoLGA – Voice over LTE Generic Access

VoLTE – Voice over LTE

VNF- Virtual Network Function

VPN – Virtual Private Network

WAP – Wireless Application Protocol

WBA – Wireless Broadband Alliance

WCDMA – Wideband Code Division Multiple Access

WCS – Wireless Communication Service

WebRTC – Web Real-Time Communication

Wi-Fi – Wireless Fidelity

WiMAX – Worldwide Interoperability for Microwave Access

WLAN – Wireless Local Area Network

WMAN – Wireless Metropolitan Area Network

WMM – Wi-Fi Multimedia

WRC – World Radiocommunication Conference

Additional Information

5G Americas maintains market information, LTE deployment lists, and numerous white papers, available for free download on its web site: http://www.5gamericas.org.

If there are any questions regarding the download of this information, please call +1 425 372 8922 or e-mail Anushka Bishen, Public Relations Coordinator at info@5gamericas.org.

This white paper was written for 5G Americas by Rysavy Research (http://www.rysavy.com) and utilized a composite of statistical information from multiple resources.

G Select Language | ▼

Robert P. Duncan A.B., S.M., M.B.A., Ph.D.: Government warfare and surveillance system architect, author, and independent investigator..

Presents

ASK ME ANYTHING ON REDDIT. POST QUESTION&VOTE UP. INDEFINITE Q&A. ALSO LATEST CLASSIFIED INFO.

SEE YOU AT NOVEMBER 5th 2017 WHITE HOUSE MMM!!!! BE THERE!

Project Soul Catcher Vol. 2: Secrets of Cyber and Cybernetic Warfare Revealed (click book image to buy)

Volume 2 details the CIA's practices of interrogation and cybernetic mind control in their pursuit to weaponize neuropsychology. It covers the art of bio-communication war. Human beings are complex machines but their inner workings have been deciphered. Mind control and brainwashing have been perfected in the last 60 years. Hacking computers and hacking into individual minds are similar. The 21st century will be known as the age of spiritual machines and soulless men.

Check out Dr. Robert Duncan's other book, How to Tame a Demon: A short practical guide to organized intimidation stalking, electronic torture, and mind

control.

More information on Dr. Robert Duncan is available at drrobertduncan.com.

new SSL secure chat #targets

DRROBERTDUNCAN.COM GUEST BOOK & FORUM. PROFESS YOUR LOVE, DROP A LINE, SHARE YOUR STORY, WHATEVER.

Free eBook: The Matrix Deciphered by Dr. Robert Duncan, A.B., S.M., M.B.A., Ph.D., the book discloses secret details on government operations and gives a hint to what it's like to live like a victim targeted..

Bio of Dr. Robert Duncan, from his book The Matrix Deciphered:

"True nobility is exempt from fear".
– King Henry the Sixth, Part II (Act IV, Scene I).

Call me The Saint. I am the all American - prep school, Harvard College graduating with honors in computer science and a minor in premedical studies, and advanced degrees from Harvard and Dartmouth in business and science. My famous ancestors are President Lincoln, King Duncan of Scotland, and Governor William Bradford, the first governor of Massachusetts.

My research interests have been neural networks, virtual reality, and EEG controlled robotics. Before graduate school I worked for the Department of Defense, Navy, NATO, and various intelligence agencies computer science projects. I have done business consulting and computer consulting for the largest companies in the world. I have been a professor, inventor, artist, and writer. I am one of the last Renaissance men.

My projects have included algorithms for Echelon and CIA natural language parsing and classification of document content, IRS formula for red flagging audits, writing the artificial intelligence code to automate tracking of the Soviet Nuclear Submarine Fleet and all water vessels, work integrating HAARP with SIGINT SIGCOM and SPAWAR. I have worked on projects for the Justice Department connecting local, state, and federal databases for the tracking of terrorists. I developed a system for the FBI to track license plates past toll booths and other locations. I worked on the soldier 2000 program to create body networks for reading vital signs and other information. A system I worked on called Snyper is operational in Iraq which triangulates on intercity conflict gun shots. I have been to a couple secret bases in the so called free world. I have developed telemedicine robotic surgery and virtual reality applications for the Army. For DARPA, I have worked on satellite computer vision target tracking applications and tank simulation as well as integration of the land, sea, and air surveillance systems like SOSSUS, towed arrays, and others.

Projects that I have worked on outside of government contracts include my thesis on computer generated holography, a project making paralyzed people walk again using choreographed stimulated muscles movements, face recognition, voice identification and recognition, finger print recognition, and neural network robotic controller. My research interests moved to enhanced reality heads up displays and wearable computing systems. My current research involves finding a cure for the mind control directed energy weapons fiasco. The integrated global surveillance grid is actually part of the holy grail of weapons and human control systems.

My apologies to the human race for my contribution to tyranny. I was tricked into thinking it would not be misused by corrupt government especially in my beloved country. I was wrong. The Joint Non-Lethal Weapons Directorate has Skip Green on the governing board. One of my old colleagues at a technology think tank in Cambridge partly in charge of the radio frequency weapons testing for neurological disruption now torturing and killing people worldwide. Several other US Navy and UK Navy scientists have been knocked-off and that is why I have come forward. I know my time maybe near. I am currently a professor in computer science and business, helping to educate the public on government corruption, greed, and stupidity. Like my ancestor Lincoln, I am fighting against mental slavery in a new Civil War engineered by the same useless elements in over 80 government agencies who have tested radiological, drug, and viral weapons on unsuspecting Americans for over 45 years.

I have great pride in the fundamental and constitutional values of America and mean no disrespect by my blasting the incompetence, apathy, and stupidity of those involved in the conspiracy of involuntary biological, chemical, and psychotronic human effects testing. There are so many brave men and women serving in the armed forces who fight and protect us for the American values they believe in. But because of the silos of information called compartmentalization in the security agencies and the lack of accountability and oversight of black operations and some top secret projects, diabolical elements become rogue or worse destroying the very country they are tasked with protecting turning it into a hypocracy rather than a democracy.

My goal is to awaken Americans to the continued horrific acts of military and CIA weapons testers in this country and the other branches of government's inability to stop them and hold them accountable.
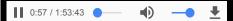
All I ask from the reader is to listen to testimony and validate facts presented here and come to their own conclusion. Then fight to win back America from this silent overthrow. This has been my project for about two years, investigating government corruption, incompetence, and cover-up upon which I accidentally stumbled one day while looking at a reverse MRI scheme to inject electromagnetic signals into the brain for virtual reality applications. I have interviewed over two hundred people and worked on the highest level of military projects for the U.S., NATO, and U.K. and have given videotaped testimony to senators and representatives on this topic resulting in lip service since they have no real power to enforce. Two high level FBI agents and a couple CIA agents have come forward to validate the existence of a MKULTRA like project that continues to grab random people for mind control experimentation. Two of these have since become part of the program and endure daily psychotronic tortures. All the torture can be done using directed energy psychotronic weapons with the so called continental ballistic missile surveillance defense grids.

I apologize to the human race for any contribution to these 4th generation weapons that I may have worked on that are more horrific than the nuclear bomb and whose cover-up is more pervasive than the Manhattan Project. And because of the horrific acts of violence being committed on as many as two thousand Americans as far as my research has uncovered and many others in other countries, I understand the extreme risk to my own welfare that publishing this material will have. But freedom is so important to me that I know full heartedly that the human race

must have an open discussion on these weapons to decide their own fate before the point of decision is gone, that I am willing to risk divulging so called national secrets. All I offer you is the truth.

All information presented in this book was received through legitimate channels such as the Freedom of Information Act, military documents, victim testimony, and turned agents. I still hold valid the oath I took to keep secret the details of the projects that I worked on under DoD budgets. The majority of the proceeds of this book will go towards helping the psychotronic experiment survivors and surviving families of those that have perished. When the government fails, business and the citizens must look out for each other.

Dr. Robert Duncan in Energy Weapons & Testing, Coast to Coast AM broadcast December 5th 2006 (2.75+ million listeners weekly, syndicated in most cities and areas of the United States, and other places around the world.)..

⏸ 0:57 / 1:53:43   ━━●━━━━━   🔊 ━●━━━   ⬇

Download - Dr. Robert Duncan, Coast to Coast AM interview - Energy Weapons & Testing.

Energy Weapons & Testing synopsis:

Independent investigator Robert Duncan discussed directed energy and neurological weapons and his contention that they've been tested on the public at large. While directed energy is used in microwaves, to remove kidney stones, and in non-invasive surgery, it's also been developed extensively for military purposes, he reported.

The civilian population was targeted for experiments, in programs such as MK-ULTRA, starting after WW II, when Nazi scientists were imported to the US, some working on scalar or gravity weapons, said Duncan. The town of Taos, NM, where a hum was heard by many citizens, was the subject of a directed energy experiment by the U.S. Navy, he claimed. And most recently the Active Denial System (see article below) was tested on human subjects.

Duncan said he interviewed over 600 mind control victims (Dec 5th 2006 report), and found some validity to their allegations. There are weapons that can project voices into people's heads such as one system known as "The Voice of God," he detailed. Blocking techniques include jamming the signals with electronic scramblers, and using shields with metal alloys or mylar. He recommended the following websites for further information/assistance: mindjustice.org, raven1.net, freedomfchs.com.

The Goodbye Weapon

A new non-lethal weapon, the Air Force's Active Denial System, or ADS, has been certified for use in Iraq, after extensive testing. The ADS shoots a beam of waves that causes extreme (but temporary) pain and induces what experimenters call the "Goodbye effect," or "prompt and highly motivated escape behavior." Wired News obtained documents about the weapon and has published two reports (article(1)/documentation(2)). Additionally, the report's author David Hambling has posted a commentary(3) on the subject at Defense Tech.

1. http://www.wired.com/news/technology/0,72134-0.html
2. http://www.wired.com/news/technology/0,72236-0.html
3. http://www.defensetech.org/archives/003036.html

Coast to Coast AM guest bio:

Dr. Robert Duncan holds multiple degrees from Harvard University and Dartmouth College amongst others. He has had the most expensive American education money can buy. He is an investigator, author, and soon be movie producer on the topics of directed energy, neurological weapons, psychological, and information warfare. His movie is called "The Enemy Within - Psychic Warfare". A book he is co-authoring will be out in a few months called "Hacking the Human Mind".

Dr. Duncan has worked as a business and information technology consultant to the Fortune 500. He has worked for companies like Oracle Corporation, BEA systems, HP, BBN, and as a professor at a college. For the department of defense he has written the artificial intelligence code to track the Soviet nuclear submarine fleet with passive and active acoustical arrays and has been to a couple secret NATA Navy underground bases in Europe.

-Todd Giffen
Target since birth as I was born into American enslavement, surveillance, and control, under attack with directed energy since 2008
503-967-5202
case@oregonstatehospital.net
obamasweapon.com

"The Use of Psychotronic Weapons"
"What everyone should know"

(courtesy : http://www.mindjustice.org/cherkova.htm)

Definition:

Definition of psychotronic (psycho-physical) weapons.

Psychotronic Weapons (PF- weapons) this is the totality of all possible methods and means (techno-genic, suggestive, pharmacological, paranormal, complexes, and others) of hidden, forced influences on the psyche of a person for the purpose of modifying his consciousness, behavior and health for what is desired in the way of influencing aspects of control..."

Second definition of psychotronic weapons (as an epigraph): "Applications in military and for antisocial purposes are possibilities also for knowledge of psychotronics, its means, methods, systems, 'generators', and there is the path for the transformation of the humanitarian, essential attributes of life and society in the weapons. These are also psychotronic weapons. This is not only dangerous, this is deadly!"

Types of Weapons:

- Electromagnetic Weapons
- Microwave Weapons
- Non-Lethal Weapons
- ELF (Extremely Low Frequency) Weapons
- Directed Energy Weapons
- Acoustic Weapons
- Psychotronic Weapons
- RF (Radio Frequency) Weapons
- Soft Kill Weapons
- Less-Than-Lethal Weapons

How it affects (Physical & psycho-physical damages):

Physical damage, visible results achieved at the moment of influence.

* burns, pricks, cuts, wounds (ray weapons, laser)
* bleeding into the brain (laser ray with energy pulses of 40 joules) and on the skin energy in pulses: 3 - 100 joules)
* Blindness in the eyes, loping of eyelashes (laser ray)
* loss of memory (influence of frequency of up to 100Hz. and a reduction to 5 Hz.)
* destruction of the vestibulary [inner ear] apparatus, spatial orientation
* symptoms of [***] [with a large exposure and strength, paralysis (UBCh) in the area of the beginning of the spinal column
* irregular sleeping, heart-beating
* numbness of the hands; at night, cramping of the legs
* sharp decrease in hemoglobin, micro-burns to the skin, ionization of the blood and body fluids, an increase in the deterioration rate of clothing material, photographic material [illuminating], (thermal-electronic emissions from lamps in the room with the feeding units of electrical networks and fittings of the wall, [...]
* vibrations (separate organs can vibrate: the heart, kidneys, etc.), [...] (from

* vibrational technologies)
* pressure, coolness of the skin
* feelings of non-specific fear and panic (from the weapons) [...] tremors, tapping, destruction of fragile objects from the waves, from the impulses and from other manifestations of technological poltergeist

Psycho-physical damage depending on intensity and duration of effects. Damage in the physical sphere (up to programmed premature death).

Stage I.

* changes in the blood (early signs)
* overheating substances especially and increase in crystallization, leading to cataracts (early sign)
* ill reactions along the path of the bio-energetic potential from the effects [...]
* prickly sensations, tremors to the muscles of the feet
* aggravation of chronic illnesses
* internal bleeding (nose, stomach, gynecological)
* auditory tones of various frequencies
* sharp pain in the internal ear
* dull pains in the area behind the ears
* internal trembling, vibrations
* head pain in the temple and a warmness in that area, pain in the eye-sockets
* likelihood of miscarriage, an increase of 80% for chances of premature births, anomalous
* development of the fetus, often with defective brains

Stage II.

* "radio-sound" effect (before this is long-conducted the resonance-loosening of spinal and brain fluids at the frequency of 2-10 Hz. for maintaining the fluids of theskull and brain in stimulation
* steady reactions to all aspects of coding
* lowering of energy (observed with irradiation at the frequency carrying frequencies of ~780 Hz.and amplitude modulation ~71 Hz.)
* aggravation of dermatosis
* eye pains (cataracts, asymmetry in lenses of the eyes, puffiness [?] of the retinas)
* aggravation of the course of chronic illnesses
* destruction of the functions of the kidneys, liver, organs of digestion, circulatory system, bleeding in gingivae, gums, peridontal and others
* gauntness
* "sand" in the eyes, burning sensation, loss of hair, brittleness of nails
* compression pain in the heart fits of coughing
* changes of the color of the face (bronze shading of the skin)
* asymmetrical puffiness of the face
* disfiguration of the face, the intentional emaciation of separate muscles, cartilage
* syndromes of destruction to the external nerves of the thighs
* impotence

Stage III.

* destruction of the brain, irreversible changes to the structure of the brain (academician V.P. Vekhtereva)

* irreversible processes in the muscle material, in the bones of the skull, and in the structures of neurons "..minimalization of life" --external reaction to EMP appears as an illness, up to heavy, sharp, chronic, organic pathological processes (academician V.P. Kaznacheev)
* "..the entire speeding up of the aging process, sclerosis, oncogenesis (academician V.P. Kaznacheev)
* destruction of bio-chemical processes in the cells with distrophic changes in the organs (academician V.P. Kaznacheev)
* destruction of the cell membranes, chemical connections in the cells -leading to leukemia, cancer of the brain and prostate gland, breast cancer, (even with weak EMP, but of long duration "...the genetic apparatus is altered..." (academician V.P. Kaznacheev)
* the reflex activities are destroyed
* the functioning of the hormonal system is destroyed
* [....] Parkinson's disease (trembling of the hands) and Alzheimer's disease (from changes in the production of the hormone melanin ([...] gland) that causes degenerative changes in substances at the molecular level disfunctions of the sino-atrial node of the heart (even with a low frequency)
* infarction, cerebral thrombosis, osteo-[***], peridontal, sugar diabetes
* allergies
* epilepsy (especially caused by the multiple-rapid stimulation by a weak, low frequency waves)
* hypertension, hypotension
* programmed or instantaneous (with the subjection to extreme effects -up to the border of deadly conditions) premature death --it is well known by the experimentors that cases of death do occur (academician V.P. Kaznacheev)
* a decrease and dying off of the population, humanity, mankind

Damage in the mental sphere (up to the point of insanity).
Stage I.

* manifestations of depression
* lowering in the interest in perceived information
* shortness of temper (irritability)

Stage II.

* steady depression
* lowering of abilities for analytical thinking, some worsening of the memory
* symptoms of chronic fatigue
* superficial nature of sleep
* higher incidents of conflicts

Stage III.

* confusion in the accounting of events in real life
* loss of ability for analytical thinking
* lowering and a loss of memory
* feelings of the lack of sleep
* indifference to personal external appearance [...]
* lowering of self-control from the deficit of the hormone seratonin due to SVCh irradiation
* suicide (as the experimentors "practice" provocation techniques)
* "...invisible weak fields ...their ignoring that they are carrying with them a deadly danger"...
"...people simply start going crazy and nobody understands why..." (academician V.P. Kaznacheev)

"Consequences of effects of micro-leptonic generators are described by the leading developer in this area [from the] V. Lab. of Micro-leptonic Technologies D. Okhatryym: in 1982 there began extremely unpleasant things with the health our co-workers. At first, there arose their feeling unwell, then the loss of logic, then after this, the loss of spatial orientation. Finally, a breakdown of diseased [affected] organs. The coming out of this condition -and this includes myself -was achieved with great difficulty."

(in the book by Ju. Vorob'evskii" Knock at the Golden Gate, Moscow, 1999.

Manipulation with the memory.

* Erasure
* Blockage
* Recording of information needed by the experimentor
* training -"prompting" (after erasure)
* comparison,
* reproduction of picture a from a prior testing as though bifurcated [?]
* a shifting of the effective frequency up to 100 Hz. or a lowering to 5 Hz., with memory loss

Control of Stimulation in the brain.

* by states of consciousness (up to the loss of these)
* in sleep
* emotions
* intellectual capabilities
* attention
* speech
* activeness
* muscle tone
* stimulation of one and the same zone with stimulations of:
o 30 -40 mV- activization of attention, memory, movement;
o 60-70 mV- "emotiono-genic" effects

Different methods:

Types of Psychotronic, psychotropic and other Psy-Influences.

* radio-waves
* laser-irradiation
* infra-red irradiation
* ultra-violet irradiation UF
* x-ray irradiation
* gravitational fields
* torsional fields (hypothetical)
* leptonic and other fields

Acoustic Waves.

* infra-sonic (IZ) from 0 Hz. to 15-20 KHz.
* ultra-sonic(UZ) from 15-20 KHz. to 10**9 Hz. (ranges for IZ and UZ)

see more similar information in -Table 3

Formation of fields:

* electro-magnetic charge
* gravitational-mass
* torsional (surrounding spin fields) -spin, with any rotating body
* micro-leptonic energy (in free aspect) -[over-abundance of stable particles]?

Operator is secretly brought in, or in a secret installation nearby

BG: Biogenerator (could be tomography)
A: Antenna (irradiation and reception)
D: operator for the gathering of information, transmission and control
ZU: reminding sytem on a magnetic strip, holographic
BIP: bio-informational [...]
KC: channel connections (IK, radio, telephone, telegraph)
E: screen display
O: operator
SP: light pen with which the operator indicates on the screen, the point of influence

Stress systems and methods of harassment.

* slander
* gestures, hand movements, expressions, tests, practical "jokes"
* taking off what is needed for the eyes (vision) [?]
* psychological pressure through family, neighbors or surrounding people, vehicle drivers
* discrediting: along 5 directions (often with the involvement of the police)
o irresponsibility
o spy
o drug dealer
o vagrant
o accomplice in a crime (in an alleged crime)
* torture, assault
* blackmail, threats
* attempts on their lives with PTOs
* death to someone close or in immediate surrounding life
* murder

Stages of Zombification (Substitution of consciousness with external control).

* programming, modeling of a cybernetic double
* conducting of programs according to non-differentiation of a proper and artificial "I"-"Not-I"
* extreme influencing, stress, harassment, irradiation, special-preparations [substances], chemicals
* taking to the limits of mental and physical capacities (survivability)
* loss of will, memory
* shutting off of his own consciousness
* replacement of consciousness with the control from his cybernetic twin
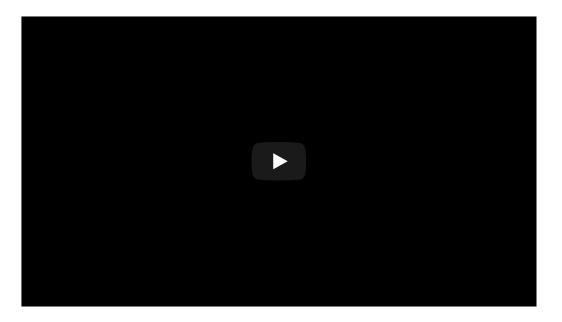
ACQUISITION, AIR, INTEL & CYBER

Pentagon Studies Weapons That Can Read Users' Mind

By SYDNEY J. FREEDBERG JR.
on July 14, 2017 at 3:15 PM
    15 Comments

28 SHARES   𝕏   f   G+   in   ✉

NEWSEUM: The troops of tomorrow may be able to pull the trigger using only their minds. As artificially intelligent underlined drones, hacking, jamming, and missiles accelerate the pace of combat, some of the military's leading scientists are studying how mere humans can keep up with the incredible speed of cyber warfare, missiles and other threats.

One option: Bypass crude physical controls — triggers, throttles, keyboards — and plug the computer directly into the human brain. In one DARPA experiment, a quadriplegic first controlled an artificial limb and then flew a flight simulator. Future systems might monitor the users' nervous system and compensate for stress, fatigue, or injury. Is this the path to what the Pentagon calls human-machine teaming?

This is an unnerving scenario for those humans, like Stephen Hawking, who mistrust artificial intelligence. If your nightmare scenario is robots getting out of control, "let's teach them to read our minds!" is probably *not* your preferred solution. It sounds more like the beginning of a movie where cyborg Arnold Schwarzenegger goes back in time to kill someone.

Navy X-47B drone.

But the Pentagon officials who talked up this research yesterday at Defense One's annual tech conference emphasized the objective was to *improve* human control over artificial intelligence. Teaching AI to monitor its user's level of stress, exhaustion, distraction, and so on helps the machine adapt itself to better serve the human — instead of the other way around. Teaching AI to instantly detect its user's *intention* to give a command, instead of requiring a relatively laborious push of a button, helps the human keep control — instead of having to let the AI off the leash because no human can keep up with it.

Official Defense Department policy, as then-Secretary Ash Carter put it, is that the US will "never" allow an artificial intelligence to decide for itself whether or not to kill a human being. However, no less a figure than the Carter's undersecretary of acquisition and technology, Frank Kendall, fretted publicly that making our robots wait for human permission would slow them down so much that enemy AI without such constraints would beat us. Vice-Chairman of the Joint Chiefs, Gen. Paul Selva, calls this the "Terminator Conundrum." Neuroscience suggests a way out of this dilemma: Instead of slowing the AIs down, make the humans' orders come faster.

DARPA's Revolutionizing Prosthetics program is devising new kinds of artificial limbs — and new ways to control them.

Accelerate Humanity

"We will continue to have humans on the loop, we will have human input in decisions, but the way we go about that is going to have to shift, just to cope with the speed and the capabilities that autonomous systems bring," said Dr. James Christensen, portfolio manager at the Air Force Research Laboratory's 711th Human Performance Wing. "The decision cycle with these systems is going to be so fast that they have to be sensitive to and responsive to the state of the individual (operator's) *intent*, as much as overt actions and control inputs that human's providing."

In other words, instead of the weapon system responding to the human operator *physically* touching a control, have it respond to the human's brain cells forming the *intention* to use a control. "When you start to have a direct neural interface of this type, you don't necessarily need to command and control the aircraft using the stick," said Justin Sanchez, director of DARPA's Biological Technologies Office. "You could potentially re-map your neural signatures onto the different control surfaces" — the tail, the flaps — "or maybe any other part of the aircraft" — say landing gear or weapons. "That part hasn't really been explored in a huge amount of depth yet."

Reading minds, even in this limited fashion, will require deep understanding and close monitoring of the brain, where thoughts take measurable form as electrical impulses running from neuron to neuron. "Can we develop precise neurotechnologies that can go to those circuits in the brain or the peripheral nervous system in real time?" Sanchez asked aloud. "Do we have computational systems that allow us to understand what the changes in those signals (mean)? And can we give meaningful feedback, either to the person or to the machine to help them to do their job better?"

DARPA's Revolutionizing Prosthetics program hooked up the brain of a quadriplegic — someone who could move neither arms nor legs — to a prosthetic arm, allowing the patient to control it directly with their thoughts. Then, "they said, 'I'd like to try to fly an airplane,'" Sanchez recounted. "So we created a virtual flight simulator for this person, allowed this neural interface to interface with the virtual aircraft, and that person flew."

"That was a real wake-up call for everybody involved," Sanchez said. "We didn't initially think you could do that."

Tony Stark (Robert Downey Jr.) relies on the JARVIS artificial intelligence — exquisitely adapted to his personal strengths and weaknesses — to help pilot his Iron Man suit. (Marvel Comics/Paramount Pictures)

Adapting To The Human

Applying direct neural control to real aircraft — or tanks, or ships, or network cybersecurity systems — will require a fundamental change in design philosophy. Today, said Christensen, we give the pilots tremendous information on the aircraft, then expect them to adapt to it. In the future, we could give the aircraft tremendous information on its pilots, then have it use artificial intelligence to adapt itself to *them*. The AI could customize the displays, the controls, even the mix of tasks it took on versus those it left up to the humans — all exquisitely tailored not just to the preferences of the individual operator but to his or her current mental and physical state.

When we build planes today, "they're incredible sensor platforms that collect data on the world around them and on the aircraft systems themselves, (but) at this point, very little data is being collected on the pilot," Christensen said. "The opportunity there with the technology we're trying to build now is to provide a continuous monitoring and assessment capability so that the aircraft knows the state of that individual. Are they awake, alert, conscious, fully capable of performing their role as part of this man-machine system? Are there things that the aircraft then can do? Can it unload gees (i.e. reduce g-forces)? Can it reduce the strain on the pilot?"

"This kind of ability to sense and understand to the state and the capabilities of the human is absolutely critical to the successful employment of highly automated systems," Christensen said. "The way all of our systems are architected right now, they're fixed, they're predictable, they're deterministic" — that is, any given input always produces the exact same output.

Air Force MQ-9 Reaper operators in training. Today's drones require extensive human oversight and are hardly user friendly.

Predictability has its advantages: "We can train to that, they behave in very consistent ways, it's easier to test and evaluate," Christensen said. "What we lose in that, though is the real power of highly automated systems, autonomous systems, as *learning systems*, of being able to adapt themselves. "

"That adaptation, though, it creates unpredictability," he continued. "So the human has to adapt alongside the system, and in order to do that, there has to be some mutual awareness, right, so the human has to understand what is the system doing, what is it trying to do, why is that happening; and vice versa, the system has to has some understanding of

the human's intent and also their state and capabilities."

This kind of synergy between human and artificial intelligence is what some theorists refer to as the "centaur," after the mythical creature that combined human and horse — with the human part firmly in control, but benefiting from the beast's strength and speed. The centaur concept, in turn, lies at the heart of the Pentagon's ideas of human-machine teaming and what's known as the Third Offset, which seeks to counter (offset) adversaries' advancing technology with revolutionary uses of AI.

The neuroscience here is in its infancy. But it holds the promise of a happy medium between hamstringing our robots with too-close control or letting them run rampant.

28
SHARES

🐦 f G+ in ✉

EDITORIAL STAFF

Colin Clark
Editor

Sydney J. Freedberg, Jr.
Deputy Editor

Internet of Things (IoT) in 5G Wireless Communications

7 authors, including:

Waleed Ejaz
Ryerson University
69 PUBLICATIONS **397** CITATIONS

SEE PROFILE

Muhammad Ali Imran
University of Glasgow
402 PUBLICATIONS **4,102** CITATIONS

SEE PROFILE

Minho Jo
Korea University
98 PUBLICATIONS **1,145** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

IU-ATC View project

Visible Light Communication View project

IEEE *Access*

Multidisciplinary : Rapid Review : Open Access Journal

GUEST EDITORIAL

Internet of Things (IoT) in 5G Wireless Communications

During the past decade, the Internet of Things (IoT) has revolutionized the ubiquitous computing with multitude of applications built around various types of sensors. A vast amount of activity is seen in IoT based product-lines and this activity is expected to grow in years to come with projections as high as billions of devices with on average 6-7 devices per person by year 2020. With most of the issues at device and protocol levels solved during the past decade, there is now a growing trend in integration of sensors and sensor based systems with cyber physical systems and device-to-device (D2D) communications. 5th generation wireless systems (5G) are on the horizon and IoT is taking the center stage as devices are expected to form a major portion of this 5G network paradigm. IoT technologies such as machine to machine communication complemented with intelligent data analytics are expected to drastically change landscape of various industries. The emergence of cloud computing and its extension to fog paradigm with proliferation of intelligent 'smart' devices is expected to lead further innovation in IoT. These developments excite us and form a motivation to survey existing work, design new techniques, and identify new applications of IoT. Researchers, scientists, and engineers face emerging challenges in designing IoT based systems that can efficiently be integrated with the 5G wireless communications.

We received enthusiastic response to our special issue call for papers. A total of nine high quality papers were received out of which only seven were selected after a thorough review process. Invited articles were sought from two highly cited and accomplished researchers. Dr. Mischa Dohler (Fellow IEEE, Kings College London, UK) and Dr. Mung Chiang (Fellow IEEE, Princeton University, USA).

The first article "MIMO-NOMA design for small packet transmission in the Internet of Things", written by a collaborative team of researchers Ding et al.,, opportunistic serving mechanism is designed as part of the effort in the novel paradigm of Multiple Input Multiple Output Non-Orthogonal Multiple Access transmission scheme. Under the proposed method, one user is completely served with their quality of service requirements completely taken care of whereas second user is served opportunistically under the NOMA paradigm. The main contribution of this article is a design with two sets of system parameters, precoding and power allocation coefficients, in order to ensure that the potential of NOMA can be realized even if the users' channel conditions are similar. Two types of power allocation policies are developed in this paper. One is to meet first user's QoS

requirements in the long term. For example, in order to satisfy its targeted outage probability. The other is in which to realize second user's QoS requirements instantaneously, e.g., the power allocation coefficients are designed to realize its targeted data rate for each channel realization.

Machine to machine communication has a significant role to play in emerging internet of things paradigm in years and decades to come. The emerging IoT-5G scenario extends sensor based IoT capabilities to robots, actuators and drones for distributed coordination and low-latency reliable execution of tasks at hand. In the invited work titled "Enabling the IoT machine age with 5G: Machine-type multicast services for innovative real-time applications" by Condoluci et al., core attention is focused on the end-to-end reliability, latency, and energy consumption comprising both up and downlinks for 5G-IoT communication. The authors propose the definition, design, and analysis of machine-type multicast service (MTMS). They recommend different procedures that need to be redesigned for MTMS and derive the most appropriate design drivers by analyzing different performance indicators, such as scalability, reliability, latency, and energy consumption. Overall, a very interesting read complemented by open problems and future research directions to pursue.

Security is one of the biggest challenges faced by Internet of Things. With devices becoming ubiquitous and pervasive in day to day lives necessitate reliable and secure algorithms. The third article, "Security enhancement for IoT communications exposed to eavesdroppers with uncertain locations" by Xu et al., develops a secure framework for eavesdroppers with Uncertain Locations in IoT. With the assumption that the locations of eavesdroppers change independently from hop to hop, authors derive an expression for the secrecy outage probability of the two-hop transmission, which is shown to be the upper bound of the outage probability when the locations of eavesdroppers remain unchanged. Following this expression, the end users formulate a secrecy rate maximization problem with the secrecy-outage probability constraint. The optimal rate design for codebooks and power allocation between the source and relay are derived. By studying the performance of the optimal scheme in some special cases, we obtain several insights concerning the setting of system parameters.

In the article, "Enabling massive IoT in 5G and beyond systems: PHY radio frame design considerations" by Ayesha Ijaz et al., the authors propose a flexible frame structure and design for massive Internet of Things (IoT) devices working in 5G wireless network. The authors also discussed

the interdependence of different frame design parameters, service requirements and characteristics of radio environment. Based on these interdependency, they provide guidelines for radio numerology design and elaborated on the frame design for IoT communications in 5G networks to support massive connection density of low-rate, low-power devices. The article concludes with some key research findings and challenges massive IoT in 5G wireless network.

It is estimated that in year 2020, 20 to 40 billion devices will be connected to the Internet as part of the Internet of Things. A critical bottleneck for realizing the efficient IoT is the pressure it puts on the existing communication infrastructures, requiring transfer of enormous data volumes. In the article ''CONDENSE: A reconfigurable knowledge acquisition architecture for future 5G IoT'', by Dejan Vukobratovic et al., the authors propose a architecture named 'Condense' which integrates the acquisition of IoT-generated data within the 3GPP MTC (machine type communications) systems. The proposed Condense architecture introduces a service within 3GPP MTC systems – computing linear and non-linear functions over the data generated by MTC devices This service brings about the possibility that the underlying communication infrastructure communicates only the desired function of the MTC- generated data (as required by the given application at hand), and not the raw data in its entirety. This transformational approach has the potential to dramatically reduce the pressure on the 3GPP MTC communication infrastructure. The article concludes by discussing challenges, provides insights, and identifies future research directions for implementing function computation and function decomposition within practical 3GPP MTC systems.

In the article, ''Frequency-domain oversampling for cognitive CDMA systems: Enabling robust and massive multiple access for Internet of Things'' by Su Hu et al., the authors utilize the concept of cognitive radio with dynamic non-continuous spectrum bands and code division multiple access to tackle the challenge of massive spectrum resource management in IoT. In order to suppress multiple access interference resulting from the non-orthogonality of partial available spectrum bins, carrier frequency offset, and spectrum sensing mismatch, the authors propose an enhanced receiver design that combines the frequency-domain oversampling scheme (FDO) and linear minimum mean square error (MMSE) method. The simulation results show that the cognitive-CDMA with FDO-MMSE receiver outperforms that with conventional per-user MMSE receiver in the presence of multipath fading channels, carrier frequency offset, and spectrum sensing mismatch.

In the invited article, ''A survey of client-controlled HetNets for 5G'' by Michael Wang et al., a comprehensive review is provided on spectrum of client-controlled HetNets for 5G networks: from the fully devolved distributed local control approach to the hybrid control approach where clients may make the decisions given some global information provided by the network. After giving a thorough review, the authors also provide future research directions

and recommendations for evolution of 5G heterogeneous networks as an enabler for next generation internet of things.

To conclude, the Editors would like to thank all authors who submitted their manuscripts against this Special issue. There had been some very exciting submissions though we could only accept best seven high quality submissions. The editorial staff also thanks the reviewers for their timely comments on the manuscript which helped us improve the presentation and quality of the content manifold. All the associate editors thank the Editor in Chief Dr. Michael Pecht for his assistance in making this special issue a reality. We would also like to thank the publications editor Ms. Kimberly Shumard for her continuous follow up on timelines and management of articles to ensure timely completion of the review process. We complete this note by letting the readers know to stay tuned with IEEE Access as we expect even more exciting issues in months to come.

WALEED EJAZ
Ryerson University,
Toronto, ON, Canada

ALAGAN ANPALAGAN
Ryerson University,
Toronto, ON, Canada

MUHAMMAD ALI IMRAN
University of Glasgow,
Glasgow, U.K.

MINHO JO
Korea University,
Sejong, South Korea

MUHAMMAD NAEEM
Ryerson University,
Toronto, ON, Canada

SAAD BIN QAISAR
National University of Sciences and Technology,
Islamabad, Pakistan

WEI WANG
Zhejiang University,
Hangzhou, China

WALEED EJAZ (S'12–M'14–SM'16) received the B.Sc. and M.Sc. degrees in computer engineering from the National University of Sciences and Technology, Islamabad, Pakistan, and the University of Engineering and Technology, Taxila, Pakistan, respectively, and the Ph.D. degree in information and communication engineering from Sejong University, South Korea, in 2014. He was with top engineering universities in Pakistan and Saudi Arabia as a Faculty Member. He was a Post-Doctoral Fellow with Queen's University, Kingston, Canada. He is currently a Senior Research Associate with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada. He also completed certificate courses on teaching and learning in higher education from the Chang School, Ryerson University. His current research interests include Internet of Things, energy harvesting, 5G cellular networks, and mobile cloud computing. He is currently serving as an Associate Editor of the *Canadian Journal of Electrical and Computer Engineering* and the IEEE Access. In addition, he is handling the special issues in *IET Communications*, the IEEE Access, and the *Journal of Internet Technology*.

ALAGAN ANPALAGAN (S'98–M'01–SM'04) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto. He is currently a Professor with the Department of Electrical and Computer Engineering, Ryerson University, where he directs the research group involved in radio resource management and radio access and networking areas within the WINCORE Lab. He served as an Editor of the IEEE Communications Surveys and Tutorials from 2012 to 2014, the IEEE Communications Letters from 2010 to 2013, the *Wireless Personal Communications* (Springer) from 2011 to 2013, and the EURASIP *Journal of Wireless Communications and Networking* from 2004 to 2009. He has co-authored three edited books, *Design and Deployment of Small Cell Networks* (Cambridge University Press, 2014), *Routing in Opportunistic Networks* (Springer, 2013), and *Handbook on Green Information and Communication Systems* (Academic Press, 2012).

Dr. Anpalagan is a registered Professional Engineer in ON, Canada, and a fellow of the Institution of Engineering and Technology. He served as the TPC Co-Chair of the IEEE Globecom'15: SAC Green Communication and Computing, the IEEE WPMC'12 Wireless Networks, the IEEE PIMRC'11 Cognitive Radio and Spectrum Management, and the IEEE CCECE'04/08. He served as the Vice Chair of the IEEE SIG on Green and Sustainable Networking and Computing with Cognition and Cooperation (2015–), the IEEE Canada Central Area Chair (2012–2014), the IEEE Toronto Section Chair (2006–2007), the ComSoc Toronto Chapter Chair (2004–2005), and the IEEE Canada Professional Activities Committee Chair (2009–2011). He was a recipient of the Dean's Teaching Award in 2011, the Faculty Scholastic, Research, and Creativity Award in 2010 and 2014, the Faculty Service Award in Ryerson University in 2011 and 2013. He also received the Exemplary Editor Award from the IEEE ComSoc in 2013 and an Editor-in-Chief Top10 Choice Award in the IEEE Transactions on Emerging Telecommunications Technology in 2012.

MUHAMMAD ALI IMRAN (M'03–SM'12) received the M.Sc. (Hons.) and Ph.D. degrees from the Imperial College London, U.K., in 2002 and 2007, respectively. He is currently a Professor in Communication Systems with the University of Glasgow, the Vice Dean of the Glasgow College UESTC, and the Program Director of Electrical and Electronics with Communications. From 2007 to 2016, he was an Adjunct Associate Professor with The University of Oklahoma, USA, and a Visiting Professor with the 5G Innovation Centre, University of Surrey, U.K.

He has led a number of multimillion-funded international research projects encompassing the areas of energy efficiency, fundamental performance limits, sensor networks, and self-organizing cellular networks. He also led the new physical layer work area for 5G Innovation center at Surrey. He has a global collaborative research network spanning both academia and key industrial players in the field of wireless communications. He has supervised 21 successful Ph.D. graduates and published over 200 peer-reviewed research papers, including over 20 IEEE Transaction papers. He secured first rank in his B.Sc. and a distinction in his M.Sc. degree along with an award of excellence in recognition of his academic achievements conferred by the President of Pakistan. He received the IEEE Comsoc's Fred Ellersick Award 2014, the Sentinel of Science Award 2016, and the FEPS Learning and Teaching Award 2014 and twice nominated for Tony Jean's Inspirational Teaching Award. He is a shortlisted finalist for The Wharton-QS Stars Awards 2014, the Reimagine Education Award for innovative teaching, and the VC's learning and teaching award in the University of Surrey. He is a Senior Fellow of the Higher Education Academy, U.K.

Dr. Imran has given an invited TEDx talk (2015) and over ten plenary talks, several tutorials and seminars in international conferences and other institutions. He has taught international short courses in USA and China. He is the Co-Founder of the IEEE Workshop BackNets 2015 and chaired several tracks/workshops of international conferences. He is an Associate Editor of the IEEE Communications Letters, the IEEE Access, and the *IET Communications* journal.

MINHO JO (SM'16) received the B.A. degree from the Department of Industrial Engineering, Chosun University, South Korea, in 1984, and the Ph.D. degree from the Department of Industrial and Systems Engineering, Lehigh University, USA, in 1994. He is currently a Professor with the Department of Computer Convergence Software, Korea University, Sejong, South Korea. He is one of the founders of the Samsung Electronics LCD Division. He was the Vice President of the Institute of Electronics and Information Engineers. He is currently the Vice President of the Korean Society for Internet Information. His current research interests include Internet of Things, AI/big data in Internet of Things, LET-Unlicensed, cognitive radio, mobile cloud computing, network security, heterogeneous networks in 5G, wireless energy harvesting, optimization in wireless communications, and massive MIMO. He is the Founder and an Editor-in-Chief of the *KSII Transactions on Internet and Information Systems* [SCI (ISI) and SCOPUS]. He served as an Editor of the IEEE Network. He is an Editor of the IEEE Wireless Communications, and an Associate Editor of the IEEE Internet of Things Journal, *Security and Communication Networks*, *Ad-Hoc & Sensor Wireless Networks*, and *Wireless Communications and Mobile Computing*.

MUHAMMAD NAEEM received the B.S. and M.S. degrees in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2000 and 2005, respectively, and the Ph.D. degree from Simon Fraser University, BC, Canada, in 2011. From 2000 to 2005, he was a Senior Design Engineer with Comcept (pvt.) Ltd. At the design department of Comcept (pvt) Ltd., he participated in the design and development of smart card-based GSM and CDMA pay phones. From 2012 to 2013, he was a Post-Doctoral Research Associate with the WINCORE Laboratory, Ryerson University, Toronto, ON, Canada. Since 2013, he has been an Assistant Professor with the Department of Electrical Engineering, COMSATS Institute of IT, Wah Campus, Pakistan, and a Research Associate with the WINCORE Laboratory, Ryerson University. He is also a Microsoft Certified Solution Developer. His research interests include optimization of wireless communication systems, nonconvex optimization, resource allocation in cognitive radio networks, and approximation algorithms for mixed integer programming in communication systems. He was a recipient of NSERC CGS Scholarship.

SAAD BIN QAISAR (S'04–M'09–SM'14) received the master's and Ph.D. degrees in electrical engineering from Michigan State University, USA, in 2005 and 2009, respectively. He is currently serving as an Associate Professor with the National University of Sciences and Technology, Pakistan, and also leading Bitsym, an Internet of Things and data-analytics enterprise. He has authored over 60 publications, inventor of various patents, has completed multiple national and international projects and winner of various awards, including the University Best Researcher Award in 2012 and the UNIDO Cleantech Award in 2015, to name a few. His company, Bitsym, has been named among the 50 Most Promising Start-ups for 2012 by Global Entrepreneurship Week, globally.

WEI WANG (S'08–M'10–SM'15) received the B.S. and Ph.D. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004 and 2009 respectively. From 2007 to 2008, he was a Visiting Student with the University of Michigan, Ann Arbor, USA. From 2013 to 2015, he was a Hong Kong Scholar with The Hong Kong University of Science and Technology, Hong Kong. He is currently an Associate Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He has authored over 90 international journal and conference papers in the area of wireless communications and networks. His research interests mainly focus on cognitive radio networks, green communications, and radio resource allocation for wireless networks.

Dr. Wang is the Associate Secretary-General of the Zhejiang Institute of Communications. He is the Editor of the book *Cognitive Radio Systems* (Intech, 2009) and serves as an Editor of the IEEE Access, Transactions on Emerging Telecommunications Technologies, and the *KSII Transactions on Internet and Information Systems*. He also serves as the Co-Chair or a TPC Member for several international conferences.

• • •

5G RADIO ACCESS

CAPABILITIES AND TECHNOLOGIES

The capabilities of 5G wireless access must extend far beyond previous generations of mobile communication. Examples of these capabilities include very high data rates, very low latency, ultra-high reliability, energy efficiency and extreme device densities, and will be realized by the development of LTE in combination with new radio-access technologies. Key technology components include extension to higher frequency bands, access/backhaul integration, device-to-device communication, flexible duplex, flexible spectrum usage, multi-antenna transmission, ultra-lean design, and user/control separation.

WHAT IS 5G?

5G radio access technology will be a key component of the Networked Society. It will address high traffic growth and increasing demand for high-bandwidth connectivity. It will also support massive numbers of connected devices and meet the real-time, high-reliability communication needs of mission-critical applications.

5G will provide wireless connectivity for a wide range of new applications and use cases, including wearables, smart homes, traffic safety/control, critical infrastructure, industry processes and very-high-speed media delivery. As a result, it will also accelerate the development of the Internet of Things.

The overall aim of 5G is to provide ubiquitous connectivity for any kind of device and any kind of application that may benefit from being connected.

5G networks will not be based on one specific radio-access technology. Rather, 5G is a portfolio of access and connectivity solutions addressing the demands and requirements of mobile communication beyond 2020.

The specification of 5G will include the development of a new flexible air interface, NX, which will be directed to extreme mobile broadband deployments. NX will also target high-bandwidth and high-traffic-usage scenarios, as well as new scenarios that involve mission-critical and real-time communications with extreme requirements in terms of latency and reliability.

In parallel, the development of Narrow-Band IoT (NB-IoT) in 3GPP is expected to support massive machine connectivity in wide area applications. NB-IoT will most likely be deployed in bands below 2GHz and will provide high capacity and deep coverage for enormous numbers of connected devices.

Ensuring interoperability with past generations of mobile communications has been a key principle of the ICT industry since the development of GSM and later wireless technologies within the 3GPP family of standards. In a similar manner, LTE will evolve in a way that recognizes its role in providing excellent coverage for mobile users, and 5G networks will incorporate LTE access (based on Orthogonal Frequency Division Multiplexing (OFDM)) along with new air interfaces in a transparent manner toward both the service layer and users.

Figure 1: The overall 5G wireless-access solution consisting of LTE evolution and new technology.

Around 2020, much of the available wireless coverage will continue to be provided by LTE, and it is important that operators with deployed 4G networks have the opportunity to transition some – or all – of their spectrum to newer wireless access technologies. For operators with limited spectrum resources, the possibility of introducing 5G capabilities in an interoperable way – thereby allowing legacy devices to continue to be served on a compatible carrier – is highly beneficial and, in some cases, even vital.

At the same time, the evolution of LTE to a point where it is a full member of the 5G family of air interfaces is essential, especially since initial deployment of new air interfaces may not operate in the same bands. The 5G network will enable dual-connectivity between LTE operating within bands below 6GHz and the NX air interface in bands within the range 6GHz to100GHz. NX should also allow for user-plane aggregation, i.e. joint delivery of data via LTE and NX component carriers.

This paper explains the key requirements and capabilities of 5G, along with its technology components and spectrum needs.

5G – REQUIREMENTS AND CAPABILITIES

In order to enable connectivity for a very wide range of applications with new characteristics and requirements, the capabilities of 5G wireless access must extend far beyond those of previous generations of mobile communication. These capabilities will include massive system capacity, very high data rates everywhere, very low latency, ultra-high reliability and availability, very low device cost and energy consumption, and energy-efficient networks.

MASSIVE SYSTEM CAPACITY

Traffic demands for mobile-communication systems are predicted to increase dramatically [1] [2]. To support this traffic in an affordable way, 5G networks must deliver data with much lower cost per bit compared with the networks of today. Furthermore, the increase in data consumption will result in an increased energy footprint from networks. 5G must therefore consume significantly lower energy per delivered bit than current cellular networks.

The exponential increase in connected devices, such as the deployment of billions of wirelessly connected sensors, actuators and similar devices for massive machine connectivity, will place demands on the network to support new paradigms in device and connectivity management that do not compromise security. Each device will generate or consume very small amounts of data, to the extent that they will individually, or even jointly, have limited impact on the overall traffic volume. However, the sheer number of connected devices seriously challenges the ability of the network to provision signaling and manage connections.

VERY HIGH DATA RATES EVERYWHERE

Every generation of mobile communication has been associated with higher data rates compared with the previous generation. In the past, much of the focus has been on the peak data rate that can be supported by a wireless-access technology under ideal conditions. However, a more important capability is the data rate that can actually be provided under real-life conditions in different scenarios.

> 5G should support data rates exceeding 10Gbps in specific scenarios such as indoor and dense outdoor environments.
> Data rates of several 100Mbps should generally be achievable in urban and suburban environments.
> Data rates of at least 10Mbps should be accessible almost everywhere, including sparsely-populated rural areas in both developed and developing countries.

VERY LOW LATENCY

Very low latency will be driven by the need to support new applications. Some envisioned 5G use cases, such as traffic safety and control of critical infrastructure and industry processes, may require much lower latency compared with what is possible with the mobile-communication systems of today.

To support such latency-critical applications, 5G should allow for an application end-to-end latency of 1ms or less, although application-level framing requirements and codec limitations for media may lead to higher latencies in practice. Many services will distribute computational capacity and storage close to the air interface. This will create new capabilities for real-time communication and will allow ultra-high service reliability in a variety of scenarios, ranging from entertainment to industrial process control.

ULTRA-HIGH RELIABILITY AND AVAILABILITY

In addition to very low latency, 5G should also enable connectivity with ultra-high reliability and ultra-high availability. For critical services, such as control of critical infrastructure and traffic

safety, connectivity with certain characteristics, such as a specific maximum latency, should not merely be 'typically available.' Rather, loss of connectivity and deviation from quality of service requirements must be extremely rare. For example, some industrial applications might need to guarantee successful packet delivery within 1 ms with a probability higher than 99.9999 percent.

VERY LOW DEVICE COST AND ENERGY CONSUMPTION

Low-cost, low-energy mobile devices have been a key market requirement since the early days of mobile communication. However, to enable the vision of billions of wirelessly connected sensors, actuators and similar devices, a further step has to be taken in terms of device cost and energy consumption. It should be possible for 5G devices to be available at very low cost and with a battery life of several years without recharging.

ENERGY-EFFICIENT NETWORKS

While device energy consumption has always been prioritized, energy efficiency on the network side has recently emerged as an additional KPI, for three main reasons:

> Energy efficiency is an important component in reducing operational cost, as well as a driver for better dimensioned nodes, leading to lower total cost of ownership.
> Energy efficiency enables off-grid network deployments that rely on medium-sized solar panels as power supplies, thereby enabling wireless connectivity to reach even the most remote areas.
> Energy efficiency is essential to realizing operators' ambition of providing wireless access in a sustainable and more resource-efficient way.

The importance of these factors will increase further in the 5G era, and energy efficiency will therefore be an important requirement in the design of 5G wireless access.

MACHINE-TYPE COMMUNICATION

Fundamentally, applications such as mobile telephony, mobile broadband and media delivery are about information for humans. In contrast, many of the new applications and use cases that drive the requirements and capabilities of 5G are about end-to-end communication between machines. To distinguish them from the more human-centric wireless-communication use cases, these applications are often termed machine-type communication (MTC).

Although spanning a wide range of applications, MTC applications can be divided into two main categories – massive MTC and critical MTC – depending on their characteristics and requirements.
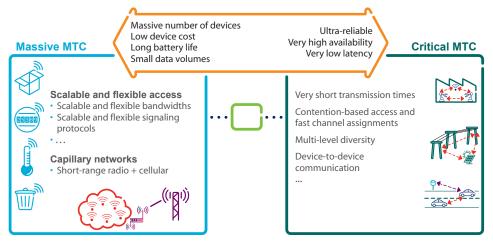
Figure 2: Massive MTC and critical MTC.

Massive MTC refers to services that typically span a very large numbers of devices, usually sensors and actuators. Sensors are extremely low cost and consume very low amounts of energy in order to sustain long battery life. Clearly, the amount of data generated by each sensor is normally very small, and very low latency is not a critical requirement. While actuators are similarly limited in cost, they will likely have varying energy footprints ranging from very low to moderate energy consumption.

Sometimes, the mobile network may be used to bridge connectivity to the device by means of capillary networks. Here, local connectivity is provided by means of a short-range radio access technology, for example Wi-Fi, Bluetooth [3] or 802.15.4/6LoWPAN [4]. Wireless connectivity beyond the local area is then provided by the mobile network via a gateway node.

Critical MTC refers to applications such as traffic safety/control, control of critical infrastructure and wireless connectivity for industrial processes. Such applications require very high reliability and availability in terms of wireless connectivity, as well as very low latency. On the other hand, low device cost and energy consumption is not as critical as for massive MTC applications. While the average volume of data transported to and from devices may not be large, wide instantaneous bandwidths are useful in being able to meet capacity and latency requirements.

There is much to gain from a network being able to handle as many different applications as possible, including mobile broadband, media delivery and a wide range of MTC applications by means of the same basic wireless-access technology and within the same spectrum. This avoids spectrum fragmentation and allows operators to offer support for new MTC services for which the business potential is inherently uncertain, without having to deploy a separate network and reassign spectrum specifically for these applications.

SPECTRUM FOR 5G

In order to support increased traffic capacity and to enable the transmission bandwidths needed to support very high data rates, 5G will extend the range of frequencies used for mobile communication. This includes new spectrum below 6GHz, as well as spectrum in higher frequency bands.

Specific candidate spectrum for mobile communication in higher frequency bands is yet to be identified by the ITU-R or by individual regulatory bodies. The mobile industry remains agnostic about particular choices, and the entire frequency range up to approximately 100GHz is under consideration at this stage, although there is significant interest in large contiguous allocations that can provide dedicated and licensed spectrum for use by multiple competing network providers.

The lower part of this frequency range, below 30GHz, is preferred from the point of view of propagation properties. At the same time, very large amounts of spectrum and the possibility of wide transmission frequency bands of the order of 1GHz or more are more likely above 30GHz.
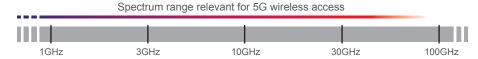
Figure 3: Spectrum relevant for 5G wireless access.

Spectrum relevant for 5G wireless access therefore ranges from below 1GHz up to approximately 100GHz, as Figure 3 shows.

It is important to understand that high frequencies, especially those above 10GHz, can only serve as a complement to lower frequency bands, and will mainly provide additional system capacity and very wide transmission bandwidths for extreme data rates in dense deployments. Spectrum allocations at lower bands will remain the backbone for mobile-communication networks in the 5G era, providing ubiquitous wide-area connectivity.

The World Radio Conference (WRC)-15 discussions have resulted in an agreement to include an agenda item for IMT-2020, the designated ITU-R qualifier for 5G, in WRC-19. The conference also reached agreement on a set of bands that will be studied for 5G, with direct applicability to NX. Many of the proposed bands are in the millimeter wave region and include:

> 24.25GHz to 27.5GHz, 37GHz to 40.5GHz, 42.5GHz to 43.5GHz, 45.5GHz to 47GHz, 47.2GHz to 50.2GHz, 50.4GHz to 52.6GHz, 66 GHz to 76GHz and 81GHz to 86GHz, which have allocations to the mobile service on a primary basis; and

> 31.8GHz to 33.4GHz, 40.5GHz to 42.5GHz and 47GHz to 47.2GHz, which may require additional allocations to the mobile service on a primary basis.

The mobile industry will strive to gain access to spectrum in the 6GHz to 20GHz range, but the policy directions being followed by regulators seem to be focused on frequency bands above 30GHz. In the US, the FCC has issued two Notices of Public Rule Making (NPRM) on bands above 24GHz. Ofcom has likewise indicated a preference for bands above 30GHz within the mobile industry.

The capacity needs of the mobile industry will continue to be served by licensed spectrum, although novel sharing arrangements for spectrum will become progressively more important as restricted opportunities for new spectrum start to impact incumbent services such as satellite communication and radio location. Two examples of sharing arrangements include LSA planned in Europe for the 2.3GHz band and the Citizens Band Radio Service for 3.5GHz in the US.

5G TECHNOLOGY COMPONENTS

Beyond extending operation to higher frequencies, there are several other key technology components relevant for the evolution to 5G wireless access. These components include access/backhaul integration, device-to-device communication, flexible duplex, flexible spectrum usage, multi-antenna transmission, ultra-lean design, and user/control separation.

ACCESS/BACKHAUL INTEGRATION

Wireless technology is already frequently used as part of the backhaul solution. Such wireless-backhaul solutions typically operate under line-of-sight conditions using proprietary radio technology in higher frequency bands, including the millimeter wave (mmW) band.

In the future, the access (base-station-to-device) link will also extend to higher frequencies. Furthermore, to support dense low-power deployments, wireless backhaul will have to extend to cover non-line-of-sight conditions, similar to access links.

In the 5G era, the wireless-access link and wireless backhaul should not therefore be seen as two separate entities with separate technical solutions. Rather, backhaul and access should be seen as an integrated wireless-access solution able to use the same basic technology and operate using a common spectrum pool. This will lead to more efficient overall spectrum utilization as well as reduced operation and management effort.

DIRECT DEVICE-TO-DEVICE COMMUNICATION

The possibility of limited direct device-to-device (D2D) communication has recently been introduced as an extension to the LTE specifications. In the 5G era, support for D2D as part of the overall wireless-access solution should be considered from the start. This includes peer-to-peer user-data communication directly between devices, but also, for example, the use of mobile devices as relays to extend network coverage.

D2D communication in the context of 5G should be an integral part of the overall wireless-access solution, rather than a stand-alone solution. Direct D2D communication can be used to offload traffic, extend capabilities and enhance the overall efficiency of the wireless-access network. Furthermore, in order to avoid uncontrolled interference to other links, direct D2D communication should be under network control. This is especially important for the case of D2D communication in licensed spectrum.

FLEXIBLE DUPLEX

Frequency Division Duplex (FDD) has been the dominating duplex arrangement since the beginning of the mobile communication era. In the 5G era, FDD will remain the main duplex scheme for lower frequency bands. However, for higher frequency bands – especially above 10GHz – targeting very dense deployments, Time Division Duplex (TDD) will play a more important role.

In very dense deployments with low-power nodes, the TDD-specific interference scenarios (direct base-station-to-base-station and device-to-device interference) will be similar to the 'normal' base-station-to-device and device-to-base-station interference that also occurs for FDD.

Furthermore, for the dynamic traffic variations expected in very dense deployments, the ability to dynamically assign transmission resources (time slots) to different transmission directions may allow more efficient utilization of the available spectrum.

To reach its full potential, 5G should therefore allow for very flexible and dynamic assignment of TDD transmission resources. This is in contrast to current TDD-based mobile technologies, including TD-LTE, for which there are restrictions on the downlink/uplink configurations, and for which there typically exist assumptions about the same configuration for neighbor cells and also between neighbor operators.

FLEXIBLE SPECTRUM USAGE

Since its inception, mobile communication has relied on spectrum licensed on a per-operator basis within a geographical area. This will remain the foundation for mobile communication in the 5G era, allowing operators to provide high-quality connectivity in a controlled-interference environment.

However, per-operator licensing of spectrum will be complemented with the possibility to share spectrum. Such sharing may be between a limited set of operators, or may occur in license-exempt scenarios. The

Citizens Band Radio Service in the US in the 3.5GHz band and the 5GHz unlicensed spectrum are examples of managed and unlicensed sharing regimes respectively.

New air interfaces like NX will likely be well served by more conventional licensed allocations of spectrum, mainly due to the need to establish a basic foundation for the technology to operate in an independent manner while interoperability is established with technologies like LTE. At some point, further allocations of spectrum for 5G may leverage the mobile industry's experience of sharing approaches in lower cellular bands.

MULTI-ANTENNA TRANSMISSION

Multi-antenna transmission already plays an important role in current generations of mobile communication and will be even more central in the 5G era, due to the physical limitations of small antennas. Path loss between a transmitter and receiver does not change as a function of frequency, as long as the effective aperture of the transmitting and receiving antennas does not change. The antenna aperture does reduce in proportion to the square of the frequency, and that reduction can be compensated by the use of higher antenna directivity. The 5G radio will employ hundreds of antenna elements to increase antenna aperture beyond what may be possible with current cellular technology.

In addition, the transmitter and receiver will use beamforming to track one another and improve energy transfer over an instantaneously configured link. Beamforming will also improve the radio environment by limiting interference to small fractions of the entire space around a transmitter and likewise limiting the impact of interference on a receiver to infrequent stochastic events. The use of beamforming will also be an important technology for lower frequencies; for example, to extend coverage and to provide higher data rates in sparse deployments.

ULTRA-LEAN DESIGN

Ultra-lean radio-access design is important to achieve high efficiency in 5G networks. The basic principle of ultra-lean design can be expressed as: minimize any transmissions not directly related to the delivery of user data. Such transmissions include signals for synchronization, network acquisition and channel estimation, as well as the broadcast of different types of system and control information.

Ultra-lean design is especially important for dense deployments with a large number of network nodes and highly variable traffic conditions. However, lean transmission is beneficial for all kinds of deployments, including macro deployments.

By enabling network nodes to enter low-energy states rapidly when there is no user-data transmission, ultra-lean design is an important component in delivering high network energy performance. Ultra-lean design will also enable higher achievable data rates by reducing interference from non-user-data-related transmissions.

USER/CONTROL SEPARATION

Another important design principle for 5G is to decouple user data and system control functionality. The latter includes the provisioning of system information; that is, the information and procedures needed for a device to access the system.

Such a decoupling will allow separate scaling of user-plane capacity and basic system control functionality. For example, user data may be delivered by a dense layer of access nodes, while system information is only provided via an overlaid macro layer on which a device also initially accesses the system.

It should be possible to extend the separation of user data delivery and system control functionality over multiple frequency bands and RATs. As an example, the system control functionality for a dense layer based on new high-frequency radio access could be provided by means of an overlaid LTE layer.

User/control separation is also an important component for future radio-access deployments relying heavily on beamforming for user data delivery. Combining ultra-lean design with a logical separation of user-plane data delivery and basic system connectivity functionality will enable a much higher degree of device-centric network optimization of the active radio links in the network. Since only the ultra-lean signals related to the system control plane need to be static, it is possible to design a system where almost everything can be dynamically optimized in real time.

An ultra-lean design combined with a system control plane logically separated from the user data delivery function also provides higher flexibility in terms of evolution of the RAT as, with such separation, the user plane can evolve while retaining system control functionality.

CONCLUSION

5G is the next step in the evolution of mobile communication and will be a key component of the Networked Society. In particular, 5G will accelerate the development of the Internet of Things. To enable connectivity for a wide range of applications and use cases, the capabilities of 5G wireless access must extend far beyond those of previous generations of mobile communications.

These capabilities include very high achievable data rates, very low latency and ultra-high reliability. Furthermore, 5G wireless access needs to support a massive increase in traffic in an affordable and sustainable way, implying a need for a dramatic reduction in the cost and energy consumption per delivered bit.

5G wireless access will be realized by the evolution of LTE for existing spectrum in combination with new radio access technologies that primarily target new spectrum. Key technology components of 5G wireless access include access/backhaul integration, device-to-device communication, flexible duplex, flexible spectrum usage, multi-antenna transmission, ultra-lean design, and user/control separation.

REFERENCES

[1] ICT-317669 METIS project, Updated scenarios, requirements and KPIs for 5G mobile and wireless system with recommendations for future investigations, Deliverable D1.5, April 2015, available at:
https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.5_v1.pdf

[2] Ericsson, Ericsson Mobility Report, November 2015, available at:
http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf

[3] Bluetooth Special Interest Group, accessed March 2016, available at:
http://bluetooth.org

[4] IETF, IPv6 over Low power WPAN (6lowpan), accessed March 2016, available at:
http://datatracker.ietf.org/wg/6lowpan

GLOSSARY

D2D	device-to-device
FDD	Frequency Division Duplex
mmW	millimeter wave
MTC	Machine-Type Communication
NPRM	Notices of Public Rule Making
OFDM	Orthogonal Frequency Division Multiplexing
TDD	Time Division Duplex
WRC	World Radio Conference

CELLULAR NETWORKS FOR MASSIVE IOT

ENABLING LOW POWER WIDE AREA APPLICATIONS

With new standards specifically targeting the connectivity requirements of Massive Internet of Things (IoT) applications, cellular networks can deliver reliable, secure and diverse IoT services using existing network infrastructure.

DELIVERING NEW VALUE IN THE NETWORKED SOCIETY

In the future, all devices that benefit from an internet connection will be connected. In this Networked Society, every person and every industry will be empowered to reach their full potential. Internet of Things (IoT) technology is a key enabler of this vision by delivering machine-to-machine (M2M) and machine-to-person communications on a massive scale.

As shown in Figure 1, Ericsson predicts there will be around 28 billion connected devices by 2021, of which more than 15 billion will be connected M2M and consumer-electronics devices [1]. A large share of these will be applications served by short-range radio technologies such as Wi-Fi and Bluetooth, while a significant proportion will be enabled by wide area networks (WANs) that are primarily facilitated by cellular networks.

THE NEW IOT LANDSCAPE

The IoT revolution offers huge potential value in terms of improved efficiency, sustainability and safety for industry and society. Analysts predict that the total added value of the IoT will be USD 1.9 trillion by 2020 [2].

The variety of applications and solutions designed for individuals, business and industry is spurring the rapid expansion of the IoT market. The IoT is playing a major role across a variety of vertical sectors, generating cost savings, new revenue streams and other benefits.

Each IoT application needs a clear value proposition and business logic in line with the prevailing ecosystem, business models and value chains of the various stakeholders. For all applications, solutions need to be integrated on platforms that can scale and handle millions of devices efficiently. Business processes for administration, provisioning and charging will have to be streamlined to minimize costs and enhance the business case.

Connected devices (billions)

Figure 1: Growth in connected devices.
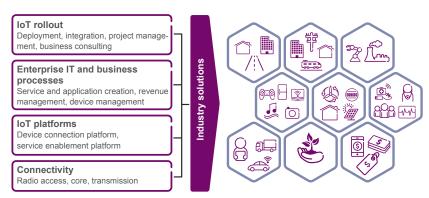
Figure 2: The new IoT landscape.

As they are largely responsible for wireless connectivity on a global scale, operators are in an excellent position to capture a share of the added value generated by the emerging IoT market. The size of this share will depend on the role that operators adopt in the value chain. This could range from being a straightforward connectivity provider (monetizing connectivity in new ways), all the way to being an end-to-end solution provider of turnkey solutions to vertical markets [3].

DIFFERENT IOT CONNECTIVITY ALTERNATIVES

Connectivity is the foundation for IoT, and the type of access required will depend on the nature of the application. Many IoT devices will be served by radio technologies that operate on unlicensed spectrum and that are designed for short-range connectivity with limited QoS and security requirements typically applicable for a home or indoor environment. Currently, there are two alternative connectivity tracks for the many IoT applications that depend on wide-area coverage:

Cellular technologies: 3GPP technologies like GSM, WCDMA, LTE and future 5G. These WANs operate on licensed spectrum and historically have primarily targeted high-quality mobile voice and data services. Now, however, they are being rapidly evolved with new functionality and the new radio access technology narrowband IoT (NB-IoT) specifically tailored to form an attractive solution for emerging low power wide area (LPWA) applications.

Unlicensed LPWA: new proprietary radio technologies, provided by, for example, SIGFOX and LoRa, have been developed and designed solely for machine-type communication (MTC) applications addressing the ultra-low-end sensor segment, with very limited demands on throughput, reliability or QoS.

One way to segment IoT applications is to categorize them according to coverage needs and performance requirements (such as data speed or latency demands).
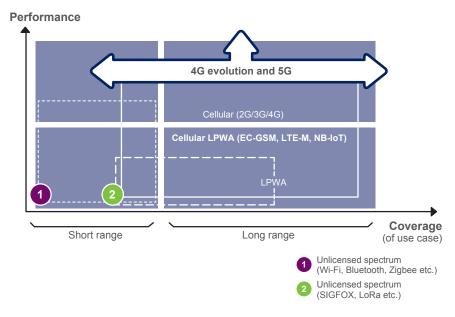
Figure 3: Technologies addressing different IoT segments.

The coverage needs of a particular use case may be highly localized (such as a stationary installation within a building), while other use cases require global service coverage (such as container tracking). 3GPP technologies already dominate use cases with large geographic coverage needs and medium- to high-performance requirements.

With new feature sets specifically tailored for LPWA IoT applications, 3GPP technologies are taking a large leap forward to cover segments with low-cost, low-performance requirements too.

CAPILLARY NETWORKS COMBINING CELLULAR AND UNLICENSED STRENGTHS

Even when existing 3GPP end-to-end connectivity is not feasible, cellular technology can still provide key benefits when used as a bridging option, i.e. as an aggregation and routing solution. This capillary network approach allows end devices to utilize varying access solutions from either the short range or LPWA domain and access the cellular networks via a gateway device. Capillary networks enable the reuse of cellular functions and assets such as security, device management, billing and QoS without requiring each end device to be cellular-enabled.

A WIDE RANGE OF IOT REQUIREMENTS

There will be a wide range of IoT use cases in the future, and the market is now expanding toward both Massive IoT deployment as well as more advanced solutions that may be categorized as Critical IoT, as shown in Figure 4.
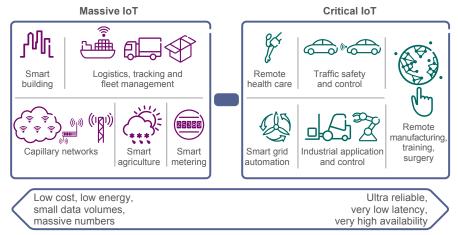
Figure 4: Differing requirements for Massive and Critical IoT applications.

At one end of the scale, in Massive IoT applications – typically sensors that report to the cloud on a regular basis – the end-to-end cost must be low enough for the business case to make sense. Here, the requirement is for low-cost devices with low energy consumption and good coverage.

At the other end of the scale, Critical IoT applications will have very high demands for reliability, availability and low latency. These use cases are enabled by LTE or 5G capabilities. Here, the volumes are typically much smaller, but the business value is significantly higher.

There are, however, many other use cases between the two extremes, which today rely on 2G, 3G or 4G connectivity.

MASSIVE IOT USE CASE DIVERSITY

The Massive IoT market segment includes several applications widely used in industries and societies, as shown in Figure 5.
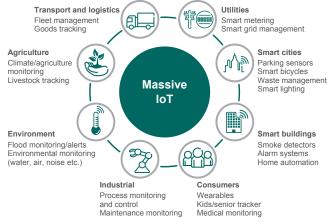
Figure 5: Industry and society applications enabled by LPWA.

The potential applications for the IoT run into the millions, with a huge variety of requirements regarding cost, battery life, coverage, connectivity performance (throughput and capacity), security and reliability. Figure 6 illustrates some such applications and their requirements regarding devices and connectivity.

Some devices will only send a few messages per day – such as status indicators for temperature – while others may need to transmit a video stream to guide a remote repair technician, for example. The difference in throughput requirements is huge. If operators or service providers handle several applications, it may be of great benefit to be able to harmonize communication modules, so that they all use the same underlying radio solution to reduce operational and fault management effort and complexity.

Many higher-value applications will require two-way communications – in other words, an uplink as well as a downlink – to enable monitoring and control of devices in systems like heating, ventilation and cooling plants. The long lifetime of many IoT applications makes it invaluable to be able to perform over-the-air device updates for new functionality or parameter settings. The amount of data sent for such updates can often be more demanding for the network than the monitoring or control application itself.

Relatively simple uplink-focused applications can benefit greatly from a bi-directional link to provide robustness. For example, a connected smoke detector must deliver a smoke alarm with absolute certainty. The ability of a network to provide acknowledgements of a received message enables better fault management and the required level of reliability. Positioning can be used to locate the sensor at the point it failed and simplify operations. For tracking applications, location information is essential.

In applications like building security, sensitive information could be reported over the air, which will require strict security. Furthermore, in the case of a break-in, it is crucial that the alarm information reaches the control center in time – making QoS and two-way communication vital.

Figure 6: Device and connectivity requirements for sample IoT use cases.

KEY CHALLENGES FOR MASSIVE IOT

The key challenges to enabling large-scale uptake of Massive IoT include:

Device cost – clearly a key enabler for high-volume, mass-market applications, enabling many of the use cases.

Battery life – many IoT devices will be battery-powered, and often the cost of replacing batteries in the field is not viable.

Coverage – deep indoor connectivity is a requirement for many applications in the utility area. Furthermore, regional (or even national or global) coverage is a prerequisite for many use cases, especially within the transport area.

Scalability – in order to enable a Massive IoT market, networks need to scale efficiently. The initial investment required for supporting a limited number of devices has to be manageable, while on the other hand, the network capacity must be easy to scale to handle thousands – or millions – of devices.

Diversity – connectivity should be able to support diverse requirements from different use cases. One network supporting everything from simple static sensors to tracking services, to applications requiring higher throughput and lower latency is essential in terms of total cost of ownership (TCO).

THE ADVANTAGES OF CELLULAR TECHNOLOGIES

Each of the technologies available for IoT connectivity has its own advantages and disadvantages. However, the range of IoT connectivity requirements – both technical and commercial – means cellular technologies can provide clear benefits across a wide variety of applications, as summarized in Figure 7.

Figure 7: Advantages of cellular technologies.

In terms of global reach, cellular networks already cover 90 percent of the world's population. WCDMA and LTE are catching up, but GSM will offer superior coverage in many markets for years to come. Cellular networks have been developed and deployed over three decades, and they will be around for the foreseeable future.

The cellular mobile industry represents a huge and mature ecosystem, incorporating chipset, device and network equipment vendors, operators, application providers and many others. The global cellular ecosystem is governed by the 3GPP standardization forum, which guarantees broad industry support for future development.

When it comes to scalability, cellular networks are built to handle massive volumes of mobile broadband traffic; the traffic from most IoT applications will be relatively small and easily absorbed. Operators are able to offer connectivity for IoT applications from the start-up phase and grow this business with low TCO and only limited additional investment and effort. Operation in licensed spectrum also provides predictable and controlled interference, which enables efficient use of the spectrum to support massive volumes of devices.

Cellular connectivity offers the diversity to serve a wide range of applications with varying requirements within one network. While competing unlicensed LPWA technologies are designed solely for very low-end MTC applications, cellular networks can address everything from Massive to Critical IoT use cases.

QoS mechanisms will be essential for many IoT applications. Cellular systems have mature QoS functionality, and this enables critical MTC applications to be handled together with traffic from sensors, voice and mobile-broadband traffic on the same carrier. QoS, along with licensed spectrum as described above, provides a foundation for long-term Service Level Agreements with a specific grade of service.

Traditionally, the security mechanisms of cellular networks have been based on a physical SIM attached to the device, referred to as a Universal Integrated Circuit Card (UICC). This has also enabled roaming between operators, which has been one of the main factors behind the huge success of mobile networks. The SIM will also be essential in future IoT applications, with SIM functionality embedded in the chipset (eUICC) or handled as a soft-SIM solution running in a trusted run-time environment of the module.

With a straightforward rollout of new software, cellular networks will be able to support the full breadth of applications, ranging from low-end use cases in the LPWA segment, to the high-end segments of in-car entertainment and video surveillance. One network connecting the whole diversifying IoT market will guarantee the lowest possible TCO as well as fast time to market.
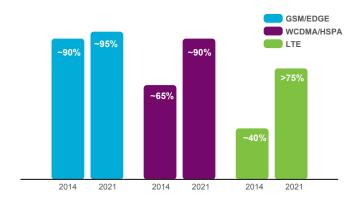
Figure 8: World population coverage by cellular 3GPP technology.

EVOLVING STANDARDS

To meet the new connectivity requirements of the emerging Massive IoT segment, 3GPP has taken evolutionary steps on both the network side and the device side.

The key improvement areas addressed in 3GPP up to Release 13 are:

> Lower device cost – cutting module cost for LTE devices by reducing peak rate, memory requirement and device complexity. The LTE module cost-reduction evolution started in Release 8 with the introduction of LTE for machine-type communication (LTE-M) Cat 1 devices with reduced peak rate to a maximum of 10Mbps, and continued in Releases 12 and 13 with reduced device complexity for lower performance and using less bandwidth or a narrowband IoT carrier to cut costs further.

> Improved battery life – more than 10 years of battery life can be achieved by introducing Power Saving Mode and/or extended discontinuous reception (eDRX) functionality. These features allow the device to contact the network – or to be contacted – on a per-need basis, meaning that it can stay in sleep mode for minutes, hours or even days.

> Improved coverage – an improvement of 15dB on LTE-M and of 20dB on NB-IoT and GSM, which translates into a seven-fold increase in the outdoor coverage area and significantly improved indoor signal penetration to reach deep indoors. This supports many IoT devices like smart meters, which are often placed in a basement.

> Support for massive numbers of IoT connections – specifically, one LTE cell site can support millions of IoT devices, depending on the use case. Core network enhancements include software upgrades for service differentiation handling, signaling optimization and high-capacity platforms (more than 30 million devices per node).
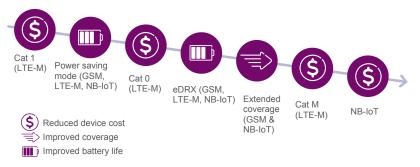
Figure 9: 3GPP evolution steps for Massive IoT.

A FULL RANGE OF CELLULAR LPWA SOLUTIONS

No single technology or solution is ideally suited to all the different potential Massive IoT applications, market situations and spectrum availability. As a result, the mobile industry is standardizing several LPWA technologies, including Extended Coverage GSM (EC-GSM), LTE-M and NB-IoT.

LTE-M, NB-IoT and EC-GSM are all superior solutions to meet Massive IoT requirements as a family of solutions, and can complement each other based on technology availability, use case requirements and deployment scenarios. LTE-M consisting of Cat 1, Cat 0 and Cat M supports a wide range of IoT applications, including those that are content-rich; NB-IoT covers ultra-low-end IoT applications with a cost and coverage advantage over LTE-M; and EC-GSM serves IoT services for all GSM markets.

For example, a smart city application such as waste management may use EC-GSM technology to provide LPWA connectivity in markets where it can be deployed on existing 2G networks; NB-IoT technology may be used for water-metering applications, which have some of the most extreme coverage requirements in underground locations. On the other hand, asset-tracking applications that can support a relatively high number of messages triggered by certain events may employ LTE-M.

EC-GSM – GLOBAL CELLULAR IOT FOR ALL GSM MARKETS

GSM is still the dominant mobile technology in many markets, and the vast majority of cellular M2M applications today use GPRS/EDGE for connectivity. GSM is likely to continue playing a key role in the IoT well into the future, due to its global coverage footprint, time to market and cost advantages.

Recognizing this – and identifying the requirements for Massive IoT discussed earlier in this paper – an initiative was undertaken in 3GPP Release 13 to further improve GSM.

The resulting EC-GSM functionality enables coverage improvements of up to 20dB with respect to GPRS on the 900MHz band.

This coverage extension is achieved for both the data and control planes by utilizing the concept of repetitions and signal combining techniques. It is handled in a dynamic manner with multiple coverage classes to ensure optimal balance between coverage and performance.

EC-GSM is achieved by defining new control and data channels mapped over legacy GSM. It allows multiplexing of new EC-GSM devices and traffic with legacy EDGE and GPRS. No new network carriers are required: new software on existing GSM networks is sufficient and provides combined capacity of up to 50,000 devices per cell on a single transceiver.

Initially part of EC-GSM but now a separate 3GPP item, eDRX improves the power efficiency – and therefore the battery life – for many use cases. eDRX improves the idle mode behavior by allowing the use of a number of inactivity timers, where the device can choose to tune in to the network and listen for downlink pages and traffic.

EC-GSM

Global solution for cellular IoT

Supported on legacy GSM equipment

Thin layer GSM

Leverages existing module ecosystem
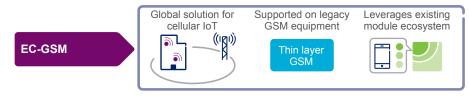
Figure 10: EC-GSM.

EC-GSM extends the data handling and power efficiency advantages that GSM/GPRS technology already offers for MTC, and it will help operators extend the service life of their huge 2G legacy base.

LTE-M – SUPPORTING A WIDE RANGE OF MASSIVE IOT USE CASES

LTE is the leading mobile broadband technology and its coverage is expanding rapidly. So far, the focus has been on meeting the huge demand for mobile data with highly capable devices that utilize new spectrum. With features like Carrier Aggregation, MIMO and Lean Carrier, the gigabit per second performance for LTE cell throughput is now reaching levels that result in an excellent mobile broadband user experience.

The advent of LTE-M signifies an important step in addressing MTC capabilities over LTE. LTE-M brings new power-saving functionality suitable for serving a variety of IoT applications; Power Saving Mode and eDRX extend battery life for LTE-M to 10 years or more. LTE-M traffic is multiplexed over a full LTE carrier, and it is therefore able to tap into the full capacity of LTE. Additionally, new functionality for substantially reduced device cost and extended coverage for LTE-M are also specified within 3GPP.
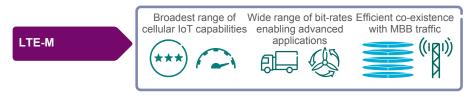
Figure 11: LTE M.

NB-IOT – SUPPORTING ULTRA-LOW-END MASSIVE IOT APPLICATIONS

In addition to LTE-M, NB-IoT technology is being standardized in time for 3GPP Release 13. NB-IoT is a self-contained carrier that can be deployed with a system bandwidth of only 200kHz, and is specifically tailored for ultra-low-end IoT applications. It is enabled using new network software on an existing LTE network, which will result in rapid time to market.

Figure 12: NB-IoT.

NB-IoT provides lean setup procedures, and a capacity evaluation indicates that each 200kHz NB-IoT carrier can support more than 200,000 subscribers. The solution can easily be scaled up by adding multiple NB-IoT carriers when needed. NB-IoT also comes with an extended coverage of up to 20dB, and battery saving features, Power Saving Mode and eDRX for more than 10 years of battery life.
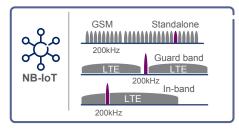
Figure 13: NB-IoT deployment.

NB-IoT is designed to be tightly integrated and interwork with LTE, which provides great deployment flexibility. The NB-IoT carrier can be deployed in the LTE guard band, embedded within a normal LTE carrier, or as a standalone carrier in, for example, GSM bands.

> Standalone deployment in a GSM low band: this is an option when LTE is deployed in a higher band and GSM is still in use, providing coverage for basic services.

> Guard band deployment, typically next to an LTE carrier: NB-IoT is designed to enable deployment in the guard band immediately adjacent to an LTE carrier, without affecting the capacity of the LTE carrier. This is particularly suitable for spectrum allocations that do not match the set of LTE system bandwidths, leaving gaps of unused spectrum next to the LTE carrier.

> Efficient in-band deployment, allowing flexible assignment of resources between LTE and NB-IoT: it will be possible for an NB-IoT carrier to time-share a resource with an existing LTE carrier. The in-band deployment also allows for highly flexible migration scenarios. For example, if the NB-IoT service is first deployed as a standalone deployment in a GSM band, it can subsequently be migrated to an in-band deployment if the GSM spectrum is re-farmed to LTE, thereby avoiding any fragmentation of the LTE carrier.

NB-IoT reduces device complexity below that of LTE-M with the potential to rival module costs of unlicensed LPWA technologies, and it will be ideal for addressing ultra-low-end applications in markets with a mature LTE installed base.

CONCLUSION

Uptake of Massive IoT is set to take off, and operators have a unique opportunity to drive the implementation of new IoT applications by offering affordable connectivity on a global scale.

For IoT applications, existing cellular networks offer distinct advantages over alternative WAN technologies, such as unlicensed LPWA. The global reach, QoS, ecosystem, TCO, scalability, diversity and security of cellular networks are all vital factors that can support the fast uptake and success of IoT. Enabled by new software in existing legacy networks, cellular networks can support a diverse range of IoT applications – ensuring the lowest possible TCO.

3GPP standardization work for GSM and LTE, and the recent addition of NB-IoT, is further improving the ability of cellular networks to address the Massive IoT market, where ultra-low end-to-end cost is a prerequisite.

GSM/GPRS, which already serves the majority of cellular-based MTC applications, is evolving with a new EC-GSM standard, which delivers significantly better energy efficiency and increased coverage. EC-GSM enhancements will cement GSM's position as a highly relevant connectivity platform for low-end, Massive IoT applications globally.

New downsized NB-IoT and LTE-M chipsets, designed for MTC, and features that improve both coverage and device battery life will boost the ability of LTE infrastructure to address the IoT market. One network that supports all applications – from advanced mobile broadband services, VoIP and all kinds of low- to high-end IoT use cases – creates a very strong value proposition.

Whether operators choose the GSM, NB-IoT or LTE-M track – or a combination of these – will depend on several factors such as technology coverage, future technology strategies and targeted market segments. Whichever path they take, they have a huge opportunity to benefit from the emerging IoT revolution. Operators can choose to continue offering telecom-grade connectivity as they do today, or they can evolve to become a platform or fully-fledged IoT service provider targeting a larger slice of future IoT revenues.

GLOSSARY

EC-GSM	Extended Coverage GSM
eDRX	extended discontinuous reception
eUICC	embedded Universal Integrated Circuit Card
IoT	Internet of Things
LPWA	low power wide area
LTE-M	LTE for machine-type communication
M2M	machine-to-machine
MIMO	multiple-input, multiple-output
MTC	machine-type communication
NB-IoT	narrowband IoT
TCO	total cost of ownership
UICC	Universal Integrated Circuit Card
WAN	wide area network

REFERENCES
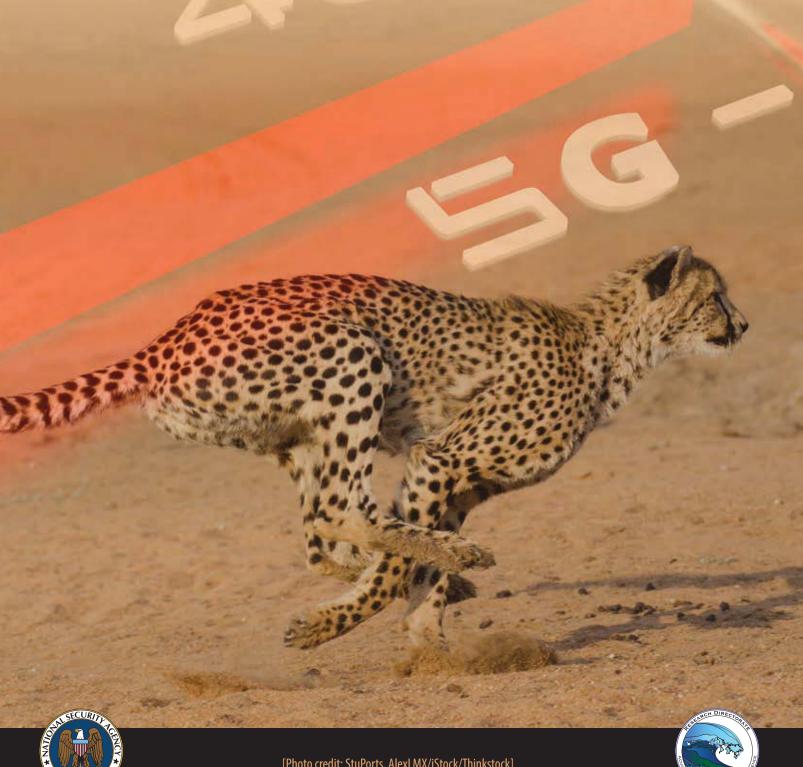
[1] Ericsson, Ericsson Mobility Report, November 2015, available at: http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf

[2] Gartner, Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020, December 2013, available at: http://www.gartner.com/newsroom/id/2636073

[3] Ericsson, White Paper: Machine-to-Machine – exploring potential operator roles, September 2014, available at: http://www.ericsson.com/res/docs/whitepapers/wp-m2m.pdf

Editor's column

While the global mobile telecommunications industry has been attempting to keep pace with ever-changing technology and consumer needs, it has not had a strong development roadmap like the International Technology Roadmap for Semiconductors (ITRS). Instead, mobile technology's development has been characterized by multinational companies pushing proposed standards for adoption by international standards bodies. These standards are subsequently adopted in a haphazard, nation-by-nation process. However, consumer desire for faster, more fully featured mobile devices has proven to be as strong a driving force as the ITRS, and has led the industry to roll out generations of new technology on a roughly 10-year basis (approximate dates: 1G - 1981, 2G - 1992, 3G - 2001, and 4G - 2009). This decade-by-decade introduction of new mobile communication technology has led to the common prediction for fifth-generation (5G) mobile to appear around 2020, in keeping with the observed linear cadence.

But why create 5G mobile anyway? The push for 5G is not just a mad rush to keep pace and provide more bandwidth to services that can already stream high-definition video. The Internet of Things (IoT; see TNW Vol. 21 No. 2) is a major driving force (among several) behind technologies being developed for 5G mobile. As personal mobile devices are more enmeshed into machine-to-machine (M2M) communications and the number of IoT sensors explodes, 5G technologies must address several needs: high-speed data rates for many more users, increased density of users, greatly increased simultaneous connections, and reduced latencies. These needs will propel many of the technologies that Dr. Farroha et al. describe in their introduction to 5G article (page 2).

One of those technologies, virtualization, has been instrumental in making efficient use of servers (virtual machine or VM) and computer networking (software-defined networking or SDN). Over the course of continuing improvement to 4G implementations, the networks are evolving into an all-digital Internet protocol packet-switched system. This evolution means that those efficiencies developed for SDN and used in

computing can be applied to mobile networking. We can see this application in more depth in the article on 5G virtualization (page 16).

Several markets have already taken advantage of 4G mobile technology, most notably media. This market is expected to expand with 5G as consumer desire for high-resolution video and augmented/virtual reality increases. Automotive, energy, health, and public safety are several areas of M2M that will be greatly enhanced by increased bandwidth and, most importantly, low-latency networks. On 13 December 2016, the US Department of Transportation proposed a rule for vehicle-to-vehicle communication and announced a vehicle-to-infrastructure communication rule to be proposed. The article on 5G and the automotive industry (page 20) provides insight into how this market segment is enhanced by 5G connectivity.

Although mobile technological developments are typically categorized into different generations, each generation covers a vast array of individual technologies and protocols that roll out as they mature. In practice, mobile service providers often work across a range of generations. This is highlighted by the fact that in many cases voice calling is handled by the 3G network while data is handled on the 4G network. Domestically, T-Mobile and AT&T did not enable voice over LTE (VoLTE) until May 2014, and did not meet the International Telecommunication Union (ITU) standard until Verizon launched LTE-Advanced in August 2016.

Many factors can impact the broad global deployment of 5G networks, such as existing technologies, geography, spectrum, and national interests. Some countries may jump to the current network generation while others may delay deployment because their current networks are considered to be "good enough." The decreased per-customer costs of updating networks in densely populated countries makes it more attractive to keep on the leading edge. The increased data rates and simultaneous connections require increased spectrum allocations, which are handled by national regulatory organizations. Several countries, most notably South Korea, have tied their technological identity to the increased

Contents

connectivity of next-generation wireless technologies. All these factors give rise to a patchwork international environment for 5G implementation.

Current forecasts for the rollout of 5G networks are in the 2020 time frame. Recent news indicates that it may be earlier than that. Samples of Qualcomm's new Snapdragon X50 4G/5G modem that uses the 27.5 gigahertz (GHz) to 28.35 GHz band—part of the spectrum opened by the Federal Communications Commission for 5G—will be available in the second half of 2017. Samsung and KT, one of South Korea's mobile providers, have announced that they will be the first to provide mobile 5G trial service at the PyeungChang 2018 Winter Olympic Games. Undoubtedly, this trial 5G service will not be fully compliant with 5G standards, which have yet to be adopted. However, we should expect things to advance quickly following the Korean introduction of 5G.

RECENT NEWS

[1] "AT&T launches first 5G business customer trial with Intel and Ericsson." AT&T. 5 Dec 2016. Available at: http://about.att.com/story/att_launches_first_5g_business_customer_trial_with_intel_and_ericsson.html.

[2] "Government must take action now to secure our connected future so we are ready for 5G, and essential services are genuinely available where they are needed." GOV.UK. 14 Dec 2016. Available at: https://www.gov.uk/government/news/government-must-take-action-now-to-secure-our-connected-future-so-we-are-ready-for-5g-and-essential-services-are-genuinely-available-where-they-are-n.

[3] Qualcomm Incorporated. "Qualcomm, Ericsson and SK Telecom announce collaboration on 5G NR trials to accelerate wide-scale 5G deployments." *PR Newswire*. 19 Dec 2016. Available at: http://www.prnewswire.com/news-releases/qualcomm-ericsson-and-sk-telecom-announce-collaboration-on-5g-nr-trials-to-accelerate-wide-scale-5g-deployments-300381570.html.

[4] Reichert C. "Huawei and NTT DoCoMo reach 11Gbps speeds in 5G Japanese field trial. 16 Nov 2016. Available at: http://www.zdnet.com/article/huawei-and-ntt-docomo-reach-11gbps-speeds-in-5g-japanese-field-trial/.

Introducing 5G

Dr. Sam Farroha, Jared S. Everett (Johns Hopkins University Applied Physics Laboratory; JHU APL),
Jason J. Uher (JHU APL), Jason R. Harper (JHU APL), Jessica K. Bridgland (JHU APL), and
Pamela M. Patton (JHU APL)

[Photo credit: Juli_Rose/iStock/Thinkstock]

The evolution of modern cellular communications has been marked by a series of technology generations. Although the technology itself tends to evolve continuously, a new generation of standards marks a revolutionary step forward, with a substantial increase in system requirements to drive fundamentally new applications. With fourth-generation (4G) networks now widely deployed, the industry has turned its sights on "the next big thing." Fifth-generation (5G) networks are expected to enable a seamlessly connected society in the time frame beyond 2020 for both people and things, including vehicles, homes, smart cities, sensor networks, and the power grid. While the Long-Term Evolution (LTE) standard will continue to evolve and play a critical role in the wireless ecosystem, 5G represents an opportunity to architect a new system that is fundamentally different without the constraint of backward compatibility with existing technologies.

What is 5G?

Although there are as yet no standards for 5G mobile networks, a number of key technology trends have emerged. This article describes seven major technology trends that will pave the way to the next generation of 5G networks.

▸ **New Flexible Radio Access Technology (RAT):** A new, non-backward-compatible RAT will be defined for 5G that is distinct from previous generations, such as 4G LTE and its evolution. New multiple access schemes under consideration include various modified Orthogonal Frequency-Division Multiplexing (OFDM)-based solutions with improved spectral efficiency. The new RAT must be flexible enough to accommodate a variety of traffic types with often conflicting radio requirements. The concept of a unified air interface has been proposed for multiplexing multiple physical layer (PHY) regions with different characteristics [e.g., transmission time interval (TTI), subcarrier spacing] on a contiguous block of spectrum [1, 2, 3]. Spectrum for the new RAT will include existing bands below 6 gigahertz (GHz), as well as new centimeter-wave (cmWave) and millimeter-wave (mmWave) bands in the 6- to 100 GHz range [4]. The new RAT must also support significantly reduced latency, with as low as 100 microsecond (μs) transmission time interval (TTI) at the PHY for the ultra-reliable and low-latency

communications (URLLC) use case [5]. Lastly, to further improve spectral efficiency, full-duplex transmission schemes have been proposed, potentially allowing the same time-frequency resources to be used for uplink and downlink transmissions simultaneously [6].

▸ **Virtualization:** Software-Defined Networking (SDN) and Network Function Virtualization (NFV) are two key architecture concepts in development to support the flexibility and mobility demands of the 5G network infrastructure [7, 8, 9]. Virtualization of network functions, which were traditionally implemented in hardware, will pave the way for commercial telecommunications operators and service providers to introduce new features and integrate new standards releases at an accelerated rate. NFV enables providers to move toward a decentralized network to increase flexibility, pushing core functions toward the edge to reduce latency, and virtualizing those functions on cloud-based servers. The proposed Cloud Radio Access Network (C-RAN) architecture, a specific use case of NFV applied to the RAN, uses a pooled architecture of baseband resources to increase scalability, physical layer flexibility, and spectral efficiency [7, 10].

▸ **Millimeter Wave (mmWave) Communications:** The term mmWave refers to carrier frequencies in the International Telecommunication Union (ITU) extremely high-frequency (EHF) band, from 30 to 300 GHz. Within the context

of 5G, the term has recently been loosely used by industry to refer to the higher frequencies from 6 to 100 GHz that are under consideration for new mobile spectrum [11, 12]. mmWave technologies are becoming an increasingly attractive solution to the problems of frequency reuse, cell density, raw data throughput, and antenna array size. This has led to a synergy between mmWave, small cell deployments, and massive multiple-input, multiple-output (MMIMO) techniques [13, 14].

▸ **Massive Multiple-Input, Multiple-Output (MMIMO) Techniques:** MMIMO is a new concept in antenna arrays that provides a number of advantages over traditional MIMO arrays currently deployed in 4G networks. Traditional MIMO arrays use only a few antenna elements (i.e., 2 to 16), whereas MMIMO uses a large number of elements in the array, currently considering a range of 128 to 512 at a minimum. Highly directional beamforming to multiple users simultaneously allows for increased user density and higher aggregate cell throughput [15]. So-called hybrid MMIMO has also been proposed; it combines beam steering with array processing techniques, such as spatial multiplexing, to increase single-user throughput [16].

▸ **Heterogeneous Networks (HetNets):** HetNets expand the mobile access network capacity by coordinating small cells with larger macro cells or offloading traffic to wireless local area network (WLAN) access points. There are two types of heterogeneity: 1) various cell sizes (e.g., macro, pico, femto) and 2) heterogeneous RATs [e.g., third-generation (3G), 4G, 5G, WLAN]. Small cells may include femto, pico, and micro cells, which can range in capacity from less than 10 to several hundred simultaneous active users. While HetNet deployments have already been introduced in 4G networks, network densification through the aggressive deployment of small cells is expected to increase significantly in future 5G networks [17, 18].

▸ **Native Machine-Type Communications (MTC) Support:** 5G networks are expected to incorporate a new model for connectivity specifically designed for MTC [19]. With the significant increase in connected machines over the last several years, a new 5G standard is seen as a prime opportunity to ensure new RATs can efficiently support a large number of connected devices with their own unique access constraints. Two categories of MTC are discussed: 1) general MTC and 2) vehicle-to-everything (V2X) MTC. General MTC devices have a few unique design and deployment considerations—namely, lower bandwidth needs, stringent power budgets, and relaxed latency requirements. V2X MTC devices, in contrast, require low-latency communications, out-of-coverage networks, and limited operation on a subscription-free basis [20].

▸ **Device-centric Architectures:** New network architectures will focus on a uniform quality of experience (QoE) for the user device, in contrast to traditional base-station-centric architectures. A number of new device-centric approaches are under consideration: decoupling the user plane and control plane, decoupling the uplink and downlink, and device-to-device communications [4, 21]. Another novel proposal is the user-centric cell or virtual cell model, which uses distributed beamforming and decoupled user/control planes to create a virtual cell around each user [2, 3]. Because the virtual cell follows the user, QoE variations are reduced and the cell-edge problem is mitigated. New device-centric architectures may significantly alter the traditional concept of cell handovers or eliminate it entirely.

Many of these technologies are already being added to the evolution of existing technologies beyond 4G, such as LTE-Advanced Pro [22].

In 2015, there was a significant increase in industry activities surrounding 5G networks. Major standards bodies, including the ITU and the Third-Generation Partnership Project (3GPP), reached important milestones in the early development of the eventual 5G standards. In September, the ITU published its vision for 5G networks [19]. The vision for International Mobile Telecommunications for 2020 and beyond (IMT-2020) defines three future-looking, high-level use cases for 5G:
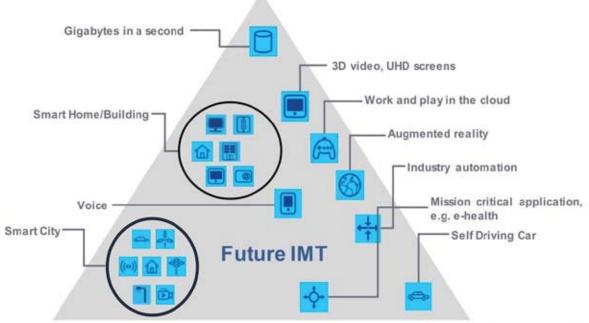
▸ **Enhanced Mobile Broadband (eMBB):** This is generally a human-centric use case driven by the exponential increase in demand for mobile access to multimedia content, services, and data.

The eMBB use case will come with new application areas and requirements that go beyond existing mobile broadband applications for improved performance and increasingly seamless user experience. This use case covers a range of scenarios, including wide-area coverage and localized high-throughput spot coverage, which will have different requirements.

▸ **Massive Machine-Type Communications (mMTC):** This use case is characterized by a large number of connected devices typically transmitting a relatively low volume of nondelay-sensitive data. Devices are intended to be low cost and have a very long battery life.

▸ **Ultra-Reliable and Low-Latency Communications (URLLC):** This use case is characterized by stringent requirements for latency, throughput, and availability. Examples include wireless control of industrial manufacturing processes, remote medical surgery, distributed smart grid automation, and transportation safety [e.g., vehicle-to-vehicle (V2V) or

vehicle-to-everything (V2X) communication]. Many companies have referred to this use case as critical MTC (cMTC) or ultra-reliable MTC (uMTC). However, based on the ITU definition in [19], this use case is not strictly limited to MTC applications.

It is important to consider that the applications that will use 5G technology do not necessarily correspond to a single use case but are more accurately described as a combination of multiple use cases. Figure 1 illustrates some examples of currently envisioned 5G applications and their relationship to these three IMT-2020 use cases [19]. Figure 2 illustrates eight key capabilities identified by ITU for IMT-2020 and their relative importance to the same three use cases [19]. Furthermore, additional future use cases are expected to emerge but cannot be accurately predicted (i.e., what will be the "killer app" in 2025?). Therefore, it is desired that 5G standards will provide the flexibility to adapt to new use cases.

FIGURE 1. 5G use cases as defined by ITU for IMT-2020 [19].
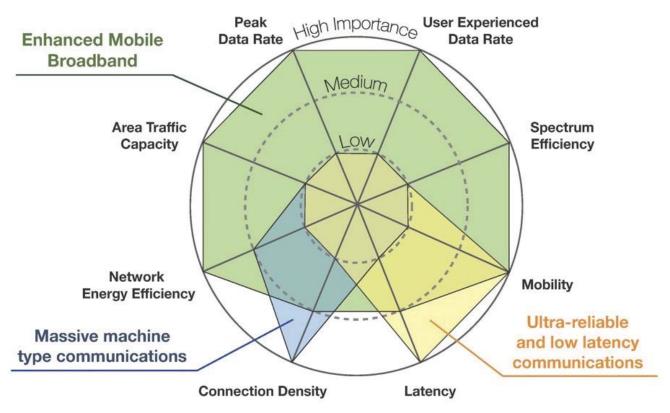
FIGURE 2. Eight key capabilities and their relative importance to 5G use cases [19].

5G Standardization

Development work toward 5G is well under way. Standards bodies are actively working on new 5G mobile technologies to be deployed in the 2020 time frame. This section summarizes the activities and corresponding 5G development timelines for three major standards bodies: ITU, 3GPP, and the Institute of Electrical and Electronics Engineers (IEEE).

Standardization in ITU

ITU is the United Nations agency responsible for promoting worldwide improvement and rational use of information and communication technology. Its members include industry, academia, and standards organizations from more than 190 member nations. The ITU Radiocommunication Sector (ITU-R) works toward worldwide consensus in the use of terrestrial and space radiocommunication services, including mobile communication technologies. Although compliance with ITU-R recommendations is not mandatory, they

nevertheless have a high degree of adoption worldwide and hold the status of international standards [23].

International Mobile Telecommunications framework
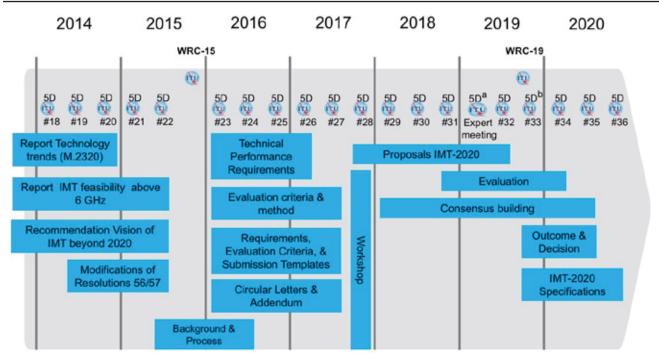
ITU-R Working Party 5D (WP 5D) is responsible for overall radio aspects of terrestrial mobile systems, referred to as International Mobile Telecommunications (IMT). The purpose of IMT is to provide high-quality mobile services with a high degree of interoperability worldwide. Since 2000, the ITU has developed the IMT standards framework in a manner that parallels cellular generations from an industry perspective. Although ITU-R WP 5D defines the requirements for IMT, it does not develop the actual radio technologies. Rather, candidate radio technologies are submitted for inclusion by external standards bodies, such as 3GPP and IEEE. For this reason, ITU-R WP 5D maintains strong cooperation with the major global standards bodies.

The first family of standards derived from the IMT concept (IMT-2000) aligned with 3G cellular. Radio technologies accepted into IMT-2000 included 3GPP Wideband Code-Division Multiple Access (WCDMA), 3GPP2 cdma2000, and IEEE 802.16 [i.e., Mobile Worldwide Wireless Interoperability for Microwave Access (WiMAX)]. The next generation of IMT standards (IMT-Advanced) aligned with 4G cellular. Radio technologies accepted into IMT-Advanced included 3GPP LTE-Advanced and IEEE 802.16m [i.e., Wireless Metropolitan Area Network (WMAN)-Advanced].

Timeline for IMT-2020

In 2012, ITU embarked on a program to develop "IMT for 2020 and beyond," setting the stage for emerging 5G research activities around the world. The program has since adopted the name IMT-2020 and forms the framework for the next generation of mobile broadband standards. The timeline for the development of IMT-2020 is shown in figure 3 [24]. The IMT 2020 timeline will essentially follow the same process used in the development of IMT-Advanced.

The IMT-2020 program is well under way, with a number of key milestones completed. In September 2015, ITU published its vision of the 5G mobile broadband connected society [19]. This document defined three high-level use cases for 5G, described earlier in this article, which have already been widely adopted by 3GPP and industry in general. In the next phase, the 2016–2017 time frame, ITU-R WP 5D will define in detail the performance requirements, evaluation criteria, and methodology for the assessment of the new IMT radio interfaces. It is anticipated that the time frame for proposals will be focused in 2018. In the 2018–2020 time frame, independent, external groups will evaluate proposals and the definition of the new radio interfaces to be included in IMT-2020 will take place. ITU-R WP 5D also plans to hold a workshop in late 2017 to discuss the performance requirements and evaluation criteria for candidate technologies for IMT-2020, as well as to provide an opportunity for presentations by potential proponents for IMT-2020 in an informal setting. The whole process is planned to be completed in 2020, when a new draft of the ITU-R recommendation with detailed

FIGURE 3. Detailed timeline and process for IMT-2020 in ITU-R [24].

specifications for the new radio technologies will be submitted for approval within ITU-R [24].

Standardization in 3GPP

3GPP is the international standards body responsible for the development and maintenance of major second-generation (2G), 3G, and 4G cellular standards. The purpose of the organization is to produce interoperable cellular communications standards, as well as studies and reports that define 3GPP technologies. The following technologies are currently maintained and evolved by 3GPP:

- ▸ Global System for Mobile Communications (GSM), General Packet Radio Service (GPRS), and Enhanced Data Rates for GSM Evolution (EDGE);
- ▸ Universal Mobile Telecommunications System (UMTS), WCDMA, High-Speed Packet Access (HSPA), and HSPA Evolution (HSPA+); and
- ▸ LTE, LTE-Advanced, and LTE-Advanced Pro.

These 3GPP technologies are constantly evolving through a series of backward-compatible releases. Since the completion of the first LTE and Evolved Packet Core (EPC) specifications, 3GPP has become the focal point for mobile systems beyond 3G. Therefore, 3GPP is expected to be a critical player in the development of 5G, and their timeline will have a direct influence on the timeline of the emerging 5G market.

3GPP is currently defining a new 5G RAT and corresponding network architecture. These are being developed within 3GPP under the working names "new radio (NR)" and "next-generation (NextGen) architecture," respectively [25]. In October 2016, 3GPP announced that the new 3GPP system will officially be known by the name "5G" from Release 15 onward [26]. Some initial standardization steps that have been taken to date include the following:

- ▸ **SMARTER study item:** In March 2015, 3GPP Technical Specification Group (TSG) System Aspects (SA) began a study item on technology enablers for new 5G services and markets, known as the SMARTER study item [27]. The objective of this study was to develop high-level use cases and identify the related high-level potential requirements to enable 3GPP network

operators to support new services and markets in 5G. Phase 1 of the SMARTER study item was completed in March 2016; results are documented in 3GPP Technical Report (TR) 22.891 to be included in Release 14 [28]. A total of 74 use cases were identified. This work prompted four building block studies that grouped the use cases into families with common requirements: massive Internet of Things (IoT), critical communications, eMBB, and network operation. The building block studies were completed in June 2016; results are documented in 3GPP TRs 22.861, 22.862, 22.863, and 22.864 to be included in Release 14 [29]. The results of the SMARTER study will form the basis for a work item to define normative stage 1 requirements for the next-generation 5G system. The work item is scheduled for completion in March 2017; results will be documented in 3GPP Technical Specification (TS) 22.261 to be included in Release 15 [29].

- ▸ **Study item on channel model for frequency spectrum above 6 GHz:** The first 5G study conducted by TSG RAN focused on developing new channel models to support high-frequency spectrum from 6 GHz to 100 GHz. The models consider a variety of scenarios including urban, rural, and indoor, as well as the impact of line-of-sight (LOS) versus non-LOS (NLOS). The study was completed in June 2016, and results are documented in 3GPP TR 38.900 to be included in Release 14 [30].

- ▸ **Study item on architecture for next-generation system:** In December 2015, 3GPP TSG SA approved a study item to design a system architecture for the next generation of mobile networks. The new architecture will support at least the new 5G RAT(s), the evolution of LTE, and non-3GPP access types and will minimize access dependencies. The study considers new approaches such as NFV and network slicing. The study item was scheduled for completion in December 2016; results will be documented in 3GPP TR 23.799 to be included in Release 14 [31].

- ▸ **Study item on scenarios and requirements for next-generation access technologies:** In December 2015, 3GPP TSG RAN approved a study item to develop deployment scenarios and requirements of next-generation access

technologies. The study identifies 12 deployment scenarios that are more diverse than those originally envisioned for legacy RATs, such as LTE and its predecessors. It also identifies key performance indicators (KPIs) and other requirements for 5G NR. The bulk of the study was completed in September 2016 to provide guidance to the ongoing technical work being performed in the RAN working groups. However, the study item will remain open until March 2017 to match the IMT-2020 timeline and ensure all IMT-2020 requirements are captured. Final results are documented in 3GPP TR 38.913 to be included in Release 14 [32, 4].

▸ **Study item on NR access technology:** In March 2016, 3GPP TSG RAN approved a study item to develop the 5G NR access technology capable of meeting the broad range of use cases defined for 5G. The study seeks to develop a single technical framework capable of addressing all

usage scenarios and requirements defined in TR 38.913 for eMBB, mMTC, and URLLC, with an emphasis on forward compatibility. The study is scheduled for completion in March 2017; results will be documented in 3GPP TR 38.912 to be included in Release 14 [1].

5G standardization activities in 3GPP will continue through 2020 and beyond, as described next.

Emerging 3GPP standardization timeline

In March 2015, 3GPP announced a tentative standardization timeline for 5G based on the ITU work plan timeline for IMT-2020 [33]. Since then, a more detailed timeline has come into focus as study items have commenced and completed, and as the 3GPP TSGs coordinate for the initial release of 5G. The timeline shown in figure 4 is a composite from multiple sources.
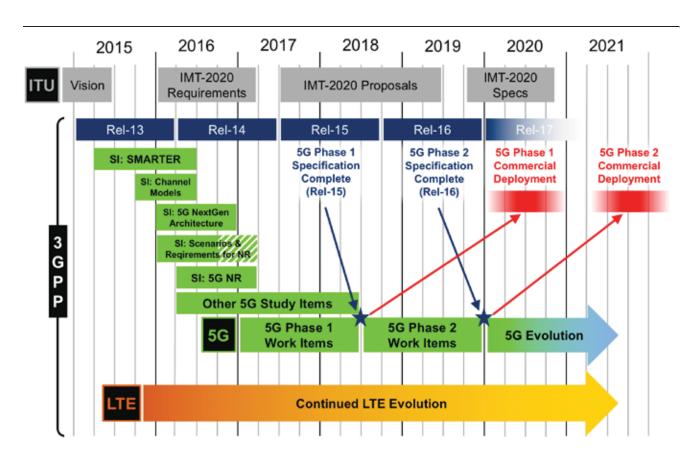
FIGURE 4. Emerging 5G standardization timeline for 3GPP. (Figure is a composite from [34] and [35] that includes additional data from various sources.)

The initial 5G study items in TSG SA and TSG RAN commenced in 2015 and 2016, as described previously. These initial 5G study items will be included in 3GPP Release 14. This work is carried out in parallel with ongoing LTE work. At the 3GPP plenary meeting in June 2016, the TSGs agreed on a work plan for the first release of 5G in 3GPP Release 15, including a clear work division between the TSGs [25].

5G work items were scheduled to begin in December 2016 for TSG SA and March 2017 for TSG RAN. The Phase 1 5G work items will fall into Release 15, with planned completion in June 2018. Additional 5G study items will continue during Release 15 in support of Phase 2. Subsequently, the Phase 2 5G work items will fall into Release 16, which will be completed around December 2019 in time for the final submission to ITU for IMT-2020. Phase 1 commercial deployments are expected to begin in 2020, followed by Phase 2 deployments in the 2021–2022 time frame. However, early pre-5G mmWave deployments may emerge in limited markets, such as South Korea or the United States, before 2020.

For example, Verizon Wireless has announced its plans to pilot a 28-GHz mmWave deployment in the United States for fixed wireless applications starting in 2017 [36]. To support this effort, the Verizon 5G Technology Forum (V5GTF)—an industry consortium led by Verizon—published an open radio interface specification in July 2016 [37]. The Verizon specification uses an OFDM-based PHY similar to time-division LTE (TD-LTE) with enhanced beamforming for operation in 28 and 39 GHz mmWave spectrum. However, with the initial focus on fixed wireless, the first release does not support user mobility. The Verizon specification can be considered pre-5G in the sense that it supports new mmWave capabilities beyond 4G but does not address all the use cases and associated requirements for 5G. The Verizon specification is expected to be incompatible with the eventual 3GPP 5G standard, potentially leading to market fragmentation [38].

3GPP phased approach to 5G standardization

3GPP TSG RAN will take a two-phased approach to developing the new 5G RAT [35]. The Phase 1 standard will define a new, non-backward-compatible 5G RAT. A subset of prioritized features and use cases will be addressed in Phase 1 to allow for early commercial deployments targeted for the year 2020. The Phase 2 standard will implement the full set of features and use cases necessary to meet the requirements for 5G. An initial proposal will be submitted to ITU as a candidate radio interface technology for IMT-2020 by the June 2019 submission deadline. The Phase 2 standard will later form the final submission around December 2019. The Phase 1 standard will be designed for forward compatibility with Phase 2 [35]. Forward compatibility means that Phase 1 must be designed from the beginning to optimally accommodate all of the features and use cases expected to be added later in Phase 2, even though those features are not yet fully implemented. Although the forward-compatibility requirement may sound straightforward, it represents a fundamental shift from the normal 3GPP standardization process, which historically has focused on a series of backward-compatible releases.

While prioritization of features between the two phases has been a topic of much debate, it is clear that the 5G Phase 1 standard will support tight interworking with LTE to simplify initial rollout. The phased approach and tight interworking with LTE means that elements of the LTE system architecture may persist in 5G deployments for some time to come. This implies that current and future work on LTE, LTE-Advanced, and LTE-Advanced Pro networks and technologies may have direct applicability to eventual 5G network deployments.

Standardization in IEEE

Initial 5G standards activities within the IEEE suggest that they do not intend to be a direct competitor with organizations like 3GPP on the radio interface between the RAN and the user equipment. Instead, IEEE has begun developing complementary technologies to support other communications requirements within the 5G ecosystem. In 2016, IEEE established two new working groups related to 5G: IEEE 1914 and IEEE 1918.

IEEE 1914 is the Next Generation Fronthaul Interface Working Group. This working group is currently developing two standards: the 1914.1 standard

for packet-based fronthaul transport networks and the 1914.3 standard for radio over Ethernet encapsulations and mappings [39, 40]. These standards focus on the fronthaul interface within the RAN between baseband units (BBUs) and remote radio heads (RRHs) to support novel RAN architectures like C-RAN, and antenna techniques like MMIMO and coordinated multi-point (CoMP) transmission and reception. The projected completion dates for these standards are August 2018 for 1914.1 and October 2017 for 1914.3.

IEEE 1918 is the Tactile Internet Working Group. This working group is currently developing the 1918.1 standard, which defines a framework for the Tactile Internet [41]. The purpose of this framework is to establish a basis for the rapid development of the Tactile Internet as a 5G and beyond application, with the expectation of additional IEEE 1918 standards to follow. The projected completion date for the 1918.1 standard is October 2018.

With respect to IMT-2020, IEEE may seek to expand the role of WLAN in 5G as a complementary radio interface for next-generation HetNets. In September 2016, the IEEE 802.11 working group sent a liaison statement to 3GPP TSG RAN and TSG SA inviting them to consider the use of IEEE 802.11-based WLAN in unlicensed spectrum as a complementary means of meeting the performance requirements of IMT-2020, potentially leading to inclusion in a joint submission to IMT-2020 [42]. This approach would be a logical extension of the increasing level of interworking between LTE and WLAN in recent standards releases. WLAN is already widely used in 3GPP networks for high data rate offloading.

Recent enhancements in radio-level interworking have increased the efficiency of these networks. Enhancements include LTE-WLAN Aggregation (LWA) and LTE WLAN Radio Level Integration with IPsec Tunnel (LWIP) in 3GPP Release 13, with further enhancements in 3GPP Release 14. Although 3GPP declined to make a decision at the September 2016 plenary meeting, the concept of a potential joint submission could represent a novel approach to IMT-2020. In contrast, previous generations of IMT saw IEEE in competition with 3GPP, with the submission of the IEEE 802.16 WiMAX family of standards to IMT-2000 and IMT-Advanced as a direct competitor in the 3G and 4G markets.

Conclusion

This article provided an introduction to major technology trends in the emergence of next-generation 5G mobile networks. These networks are expected to see initial commercial deployment starting around the year 2020. Early 5G standardization activities in the ITU, 3GPP, and IEEE were addressed.

For further information on the latest developments in 5G, the interested reader is directed to the following resources. ITU publications for IMT-2020 can be found on the IMT-2020 web page [24]. Notable documents include the ITU vision for IMT-2020 [19] as well as the technical performance requirements for IMT-2020 (scheduled for completion in February 2017). The 3GPP web site (www.3GPP.org) is the most direct source for 3GPP-related technical information. 3GPP press releases provide high-level summaries of ongoing standards activities and often include links to more detailed further reading. The latest versions of the 3GPP TR and TS documents mentioned in this article can be accessed there as well. Information on the IEEE 1914.1, 1914.3, and 1918.1 standards can be found in the approved project authorization request (PAR) documents [39, 40, 41] and the corresponding working group web pages. Lastly, the Verizon mmWave specification is available on the V5GTF website (www.5GTF.org). This pre-5G specification defines layers 1 to 3 of an open radio interface using a document structure similar to that of LTE. ↻

References

[1] 3GPP TSG RAN. "Revision of SI: Study on new radio access technology." September 2016. 3GPP Document TD RP-161596. Available at: http://www.3gpp.org/ftp/tsg_ran/TSG_RAN/ TSGR_73/Docs/RP-161596.zip.

[2] Huawei. "Vision on 5G radio access technologies." September 2015. 3GPP Document RWS 150006. Available at: ftp://ftp.3gpp.org/ workshop/2015-09-17_18_RAN_5G/Docs/RWS 150006.zip.
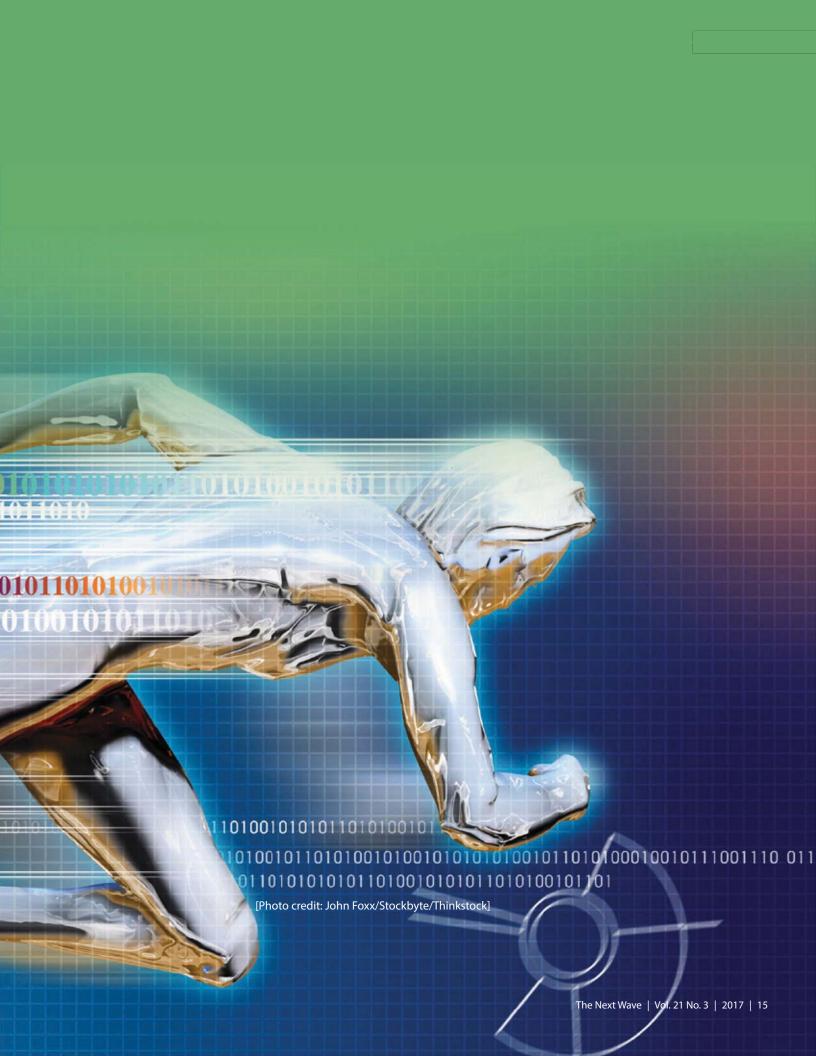
[3] ZTE. "Considerations on 5G key technologies and standardization." September 2015. 3GPP Document dRWS 150024. Available at: ftp:// ftp.3gpp.org/workshop/2015-09-17_18_RAN_5G/ Docs/RWS 150024.zip.

[4] 3GPP. "3GPP TSG RAN study on scenarios and requirements for next generation access technologies." October 2016. 3GPP Specification TR 38.913, Release 14, Version 14.0.0. Available at: http:// www.3gpp.org/DynaReport/38913.htm.

[5] Fettweis G, Alamouti S. "5G: Personal mobile internet beyond what cellular did to telephony." *IEEE Communications Magazine*. February 2014; 52(2): 140-145. doi: 10.1109/MCOM.2014.6736754.

[6] Zhang Z, Chai X, Long K, Vasilakos AV, Hanzo L. "Full duplex techniques for 5G networks: Self-interference cancellation, protocol design, and relay selection." *IEEE Communications Magazine*. May 2015; 53(5): 128-137. doi: 10.1109/ MCOM.2015.7105651.

[7] ITU-R. "Future technology trends of terrestrial IMT systems." November 2014. ITU-R Report M.2320-0. Available at: https://www.itu.int/dms_ pub/itu-r/opb/rep/R-REP-M.2320-2014-PDF-E.pdf.

[8] 4G Americas. "5G: 4G Americas' recommendations on 5G requirements and solutions." October 2014. Available at: http://www.4gamericas. org/files/2714/1471/2645/4G_Americas_ Recommendations_on_5G_Requirements_and_ Solutions_10_14_2014-FINALx.pdf.

[9] 3GPP. "3GPP NFV study: Management of mobile networks that include virtualized network functions (MANO)" [Press release]. 5 Oct 2015. Available at: http://www.3gpp.org/ news-events/3gpp-news/1738-sa5_nfv_study.

[10] Checko A, Christiansen H, Yan Y, Scolari L, Kardaras G, Berger M, Dittman L. "Cloud RAN for mobile networks—A technology overview." *IEEE Communication Surveys and Tutorials*. January 2015; 17(1): 405-426. doi: 10.1109/COMST.2014.2355255.

[11] ITU-R. "Technical feasibility of IMT in bands above 6 GHz." June 2015. ITU-R Report M.2376-0. Available at: www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2376-2015-PDF-E.pdf.

[12] Rangan S, Rappaport T, Erkip E. "Millimeter-wave cellular wireless networks: Potentials and challenges." In: *Proceedings of the IEEE;* 2014 January; 102(3). doi: 10.1109/JPROC.2014.2299397.

[13] Swindlehurst AL, Ayanoglu E, Heydari P, Capolino F. "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Communications Magazine*. September 2014; 52(9): 56-62. doi: 10.1109/MCOM.2014.6894453.

[14] Straight Path Communications. "A straight path toward 5G." September 2015. 3GPP Document RWS150087. Available at: ftp://ftp.3gpp.org/workshop/2015-09-17_18_RAN_5G/Docs/RWS150087. zip.

[15] Hoydis J, ten Brink S, Debbah M. "Massive MIMO: How many antennas do we need?" In: *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing*; 2011 Sep 27-28; Monticello, IL; p. 545–550. doi: 10.1109/ Allerton.2011.6120214.

[16] InterDigital. "Roadmap for 5G radio standardization in 3GPP." September 2015. 3GPP Document RWS 150030. Available at: ftp://ftp.3gpp.org/workshop/2015-09-17_18_RAN_5G/Docs/RWS 150030. zip.

[17] Wannstrom J, Mallinson K. "Heterogeneous networks in LTE." Available at: http://www.3gpp.org/ technologies/keywords-acronyms/1576-hetnet.

[18] Bangerter B, Talwar S, Arefi R, Stewart K. "Networks and devices for the 5G era." *IEEE Communications Magazine*. February 2014; 52(2): 90–96. doi: 10.1109/MCOM.2014.6736748.

[19] ITU-R. "IMT vision—Framework and overall objectives of the future development of IMT for 2020 and beyond." September 2015. ITU-R Recommendation M.2083-0. Available at: http:// www.itu.int/rec/R-REC-M.2083.

[20] General Motors Research. "Towards 5G: An automotive perspective." September 2015. 3GPP Document RWS 150011. Available at: ftp://ftp.3gpp. org/workshop/2015-09-17_18_RAN_5G/Docs/RWS 150011.zip.

[21] Boccardi F, Heath Jr. R, Lozano A, Marzetta T, Popovski P. "Five disruptive technology directions for 5G." *IEEE Communications Magazine*. February 2014; 52(2): 74-80. doi: 10.1109/MCOM.2014.6736746.

[22] 3GPP. "LTE-advanced pro ready to go" [Press release]. 28 Oct 2015. Available at: http://www.3gpp.org/news-events/3gpp-news/1745-lte-advanced_pro.

[23] ITU-R. "ITU-R study groups." May 2013. Available at: http://www.itu.int/dms_pub/itu-r/opb/gen/R-GEN-SGB-2013-PDF-E.pdf.

[24] ITU-R. "ITU toward 'IMT for 2020 and beyond.'" Available at: http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx.

[25] 3GPP. "3GPP on track to 5G" [Press release]. 27 Jun 2016. Available at: http://www.3gpp.org/news-events/3gpp-news/1787-ontrack_5g.

[26] 3GPP. "New partners at London 3GPP meeting" [Press release]. 24 Oct 2016. Available at: http://www.3gpp.org/news-events/3gpp-news/1800-pcg_37.

[27] 3GPP TSG SA. "New WID study on new services and markets technology enablers (FS_SMARTER) from S1-150300." March 2015. 3GPP Document SP-150142. Available at: ftp://ftp.3gpp.org/tsg_sa/TSG_SA/TSGS_67/Docs/SP-150142.zip.

[28] 3GPP. "3GPP TSG SA feasibility study on new services and markets technology enablers stage 1." June 2016. 3GPP Specification TR 22.891, Release 14, Version 14.1.0. Available at: http://www.3gpp.org/DynaReport/22891.htm.

[29] 3GPP. "SA1 completes its study into 5G requirements" [Press release]. 23 Jun 2016. Available at: http://www.3gpp.org/news-events/3gpp-news/1786-5g_reqs_sa1.

[30] 3GPP. "3GPP TSG RAN study on channel model for frequency spectrum above 6 GHz." June 2016. 3GPP Specification TR 38.900, Release 14, Version 14.0.0. Available at: http://www.3gpp.org/DynaReport/38900.htm.

[31] 3GPP TSG SA. "New WID study on architecture for next generation system." December 2015. 3GPP Document TD SP-150853. Available at: http://www.3gpp.org/ftp/tsg_sa/TSG_SA/TSGS_70/Docs/SP-150853.zip.

[32] 3GPP TSG RAN. "New study item proposal: Study on scenarios and requirements for next generation access technologies." December 2015. 3GPP Document TD RP-152257. Available at: http://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_70/Docs/RP-152257.zip.

[33] Flore D, Bertenyi B. "Tentative 3GPP timeline for 5G" [Press release]. 17 Mar 2015. Available at: http://www.3gpp.org/news-events/3gpp-news/1674-timeline_5g.

[34] Qualcomm. "5G: Views on technology and standardization." September 2015. 3GPP Document RWS 150012. Available at: ftp://ftp.3gpp.org/workshop/2015-09-17_18_RAN_5G/Docs/RWS 150012.zip.

[35] Flore D. "RAN workshop on 5G: Chairman summary." September 2015. 3GPP Document RWS 150073. Available at: ftp://ftp.3gpp.org/workshop/2015-09-17_18_RAN_5G/Docs/RWS 150073.zip.

[36] Alleven M. "Verizon's Shammo: 5G pilot in 2017 is all about fixed wireless, not mobility." *Fierce Wireless*. 2016 Apr 21. Available at: http://www.fiercewireless.com/tech/story/verizons-shammo-5g-pilot-2017-all-about-fixed-wireless-not-mobility/2016-04-21.

[37] V5GTF. "V5GTF Network and Signaling Working Group; Verizon 5G radio access; Overall description." June 2016. V5GTF Specification TS V5G.300, Release 1, Version 1.0. Available at: http://www.5gtf.org/V5G_300_v1p0.pdf.

[38] Alleven M. "Verizon's version of 5G not compatible with 3GPP's current specs—or easily upgradeable: Report." *Fierce Wireless*. 2016 Oct 28. Available at: http://www.fiercewireless.com/tech/verizon-s-version-5g-not-compatible-3gpp-s-current-specs-or-easily-upgradeable-report.

[39] IEEE. "IEEE P1914.1: Standard for packet-based fronthaul transport networks." February 2016. IEEE Approved PAR P1914.1. Available at: http://sites.ieee.org/sagroups-1914/files/2016/06/P1914.1.pdf.

[40] IEEE. "IEEE P1904.3: Standard for radio over ethernet encapsulations and mappings." June 2016. IEEE Approved PAR P1904.3. Available at: http://sites.ieee.org/sagroups-1914/files/2016/07/P1904.3.pdf.

[41] IEEE. "IEEE P1918.1: Tactile Internet: Application scenarios, definitions and terminology, architecture, functions, and technical assumptions." March 2016. IEEE Approved PAR P1918.1. Available at: http://grouper.ieee.org/groups/1918/1/files/P1918.1.pdf

[42] 3GPP TSG RAN. "IEEE 802.11 Working group liaison on the role of WLAN in IMT-2020." September 2016. 3GPP Document TD RP-161804. Available at: http://www.3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_73/Docs/RP-161804.zip.

[Photo credit: John Foxx/Stockbyte/Thinkstock]

Virtualizing the 5G Architecture

Staff Writer

Virtually changing the 5G architecture

Virtualization is set to play a major role in the evolution of the fifth-generation (5G) core network. According to industry experts, 5G will use software-centric networking technologies such as software-defined networking (SDN) and network functions virtualization (NFV), and will be natively cloud based. If correct, this will represent a major transition in system architecture and will require much greater collaboration across the networking ecosystem. The push to incorporate more cloud- or software-based components is driven by the need for greater flexibility and scalability to respond to the demands of radio access technologies that offer more bandwidth, reduced latency, and stringent quality of service (QoS) requirements. The new 5G core network must be adaptable and better equipped to handle various devices and manage capacity in near-real time. Mobile network operators see the advances in cost and efficiency that virtualization brings to other market segments and will use the emerging 5G technology to determine if these same advances can benefit the mobile market [1, 2].

SDN and NFV are two key architecture concepts in development to support the flexibility and mobility demands of the 5G network infrastructure. Virtualizing network functions that were previously implemented in hardware will allow providers to introduce new features and integrate new standards at a faster rate. SDN/NFV provides an avenue for providers to decentralize their networks, thereby increasing flexibility and reducing latency. Two areas where SDN/NFV will benefit 5G networks, and in some cases even fourth-generation (4G) mobile technology, are network slicing and cloud-radio access network (C-RAN).

Network slicing

Network slicing would promote end-to-end mobile network virtualization by "slicing" the network into virtual channels. These virtual channels would be autonomous and encompass a set of resources—physical or virtual—including bandwidth on a network link, processing capacity of servers, processing capacity of network elements, as well as operations support system (OSS) and business support system (BSS) processes. Operators could then use these channels to dynamically devote the appropriate network resources to create a "lane" in the network specifically designed for a particular use or service. This would accommodate the many use cases being put forth for 5G. The operator-led Next Generation Mobile Networks (NGMN) Alliance has sought to define categories of 5G use cases (i.e., service types) that have distinct performance characteristics and commercial potential. In a 2015 white paper, the NGMN listed eight application categories for 5G [3, 4, 5]:

1. Broadband access in dense areas,

2. Broadband access everywhere,

3. Higher user mobility,

4. Massive Internet of Things,

5. Extreme real-time communications,

6. Lifeline communications,

7. Ultra-reliable communications, and

8. Broadcast-like services.

Each of these service types demand different network requirements that are determined by the types of traffic being sent and even the types of devices sending the traffic. For example, someone downloading cat videos will not have the same bandwidth or low latency requirements as a doctor in Los Angeles performing surgery virtually on a patient in Mumbai. The end-to-end notion of network slicing could be key to 5G's ability to effectively accommodate all of these disparate use cases.
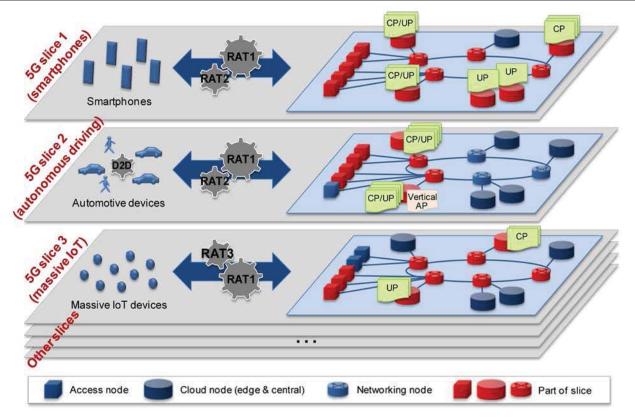
FIGURE 1. 5G's incorporation of SDN and NFV would allow network slices to be created dynamically and deployed as needed to accommodate a variety of scenarios [2].

The Open Networking Foundation, in an April 2016 white paper, called out SDN's ability to support multiple services over a common architecture as a key enabler for network slicing. SDN also allows for slices to be created dynamically and deployed as needed to accommodate a variety of scenarios (see figure 1). Currently, 4G mobile networks prioritize traffic to get a similar effect, but with more rigidity and limitations. However, as 4G networks incorporate SDN and NFV, network slicing will become an option.

C-RAN

There are more than six million base stations (excluding small cells) deployed worldwide across approximately five million different cell sites serving close to four billion users. The surge in demand for connectivity has network operators searching for ways to shrink their network footprint, lower operational expenditures (OPEX), and still meet users' demand for access. C-RAN meets these requirements and has either been implemented or trialed by several operators including Verizon, AT&T, KT (South Korea), and China Mobile. Radio base stations currently depend on special purpose-built hardware deployed at the cell site. The baseband processing unit (BBU) is the part of the RAN that is responsible for managing the radio functions (or all functions that require an antenna).

The BBU is one of the parts of the RAN that can be moved to a central location, creating a pool of BBUs to serve multiple base stations. C-RAN aims to centralize and virtualize baseband processing to reduce cell site costs and enable coordinated scheduling of resource blocks across a coverage area [6].

Figure 2 illustrates the evolution from the classic RAN model to a C-RAN setup. On the left, the classic model has the BBU deployed at the cell site connecting to the core network over IP/Ethernet transport. Any coordination between cell sites takes place over the X2 interface, which allows two sites to communicate. In a C-RAN architecture, the BBUs are pooled at a location away from the cell site. Pooling BBUs negates the need for the X2 interface as communications between cell sites now takes place internally. This is one of the reasons for the increased performance in C-RANs. The C-RAN model also makes updating the waveform and protocols easier as it only requires a software upgrade at the centralized BBU and not at each individual cell site

The move to a C-RAN architecture increases the flexibility of the network by allowing providers to, in theory, even change the types of RANs used—from 3G to 4G. For instance, in an area that has a mix of 3G and 4G users, operators can rebalance radio frequency resources by shifting more resources to 4G when 4G
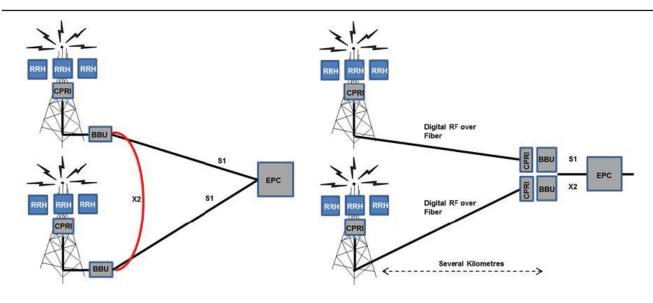
FIGURE 2. Unlike the traditional RAN model (on the left), C-RAN technology (on the right) pools the BBUs at a location away from the cell site offering an increase in performance and simpler upgrade path [6].

users are prevalent in the covered area. This shuffling of resources ensures that there is sufficient capacity for subscribers. C-RAN will also enable RAN-as-a-Service (RaaS), which will allow operators to rent RAN capacity to other operators.

Conclusion

On the face of it, virtualizing the 5G architecture seems like a necessary step towards preparing for the low latency requirements associated with 5G services. However, there are many uncertainties around 5G and a lack of visibility into what 5G will actually become through the standardization process, and at which phase different performance requirements will be supported or required. Phase II, the phase that is expected to meet International Mobile Telecommunication system for the year 2020 (IMT-2020) requirements, is expected in 2020. However, early versions or "pre-5G" offerings could be on the market before the final standard is approved. So, while these pre-5G solutions will have some 5G functionality, they will not field the full complement of improvements offered by an approved 5G system. For this reason, it is unclear to what degree virtualization techniques like C-RAN and network slicing will need to be implemented. It may come down to what the industry leaders in 5G implement in their "pre-5G" networks that will decide what a virtual 5G network will look like and when we can expect to see one. ↩

References

[1] Farroha B, Everett JS, Uher JJ, Harper JR, Bridgland JK, Patton PM. "The status of 5G and the impact on current cellular architecture," 2016. College Park (MD): Laboratory for Telecommunications Sciences and The Johns Hopkins University Applied Physics Laboratory.

[2] Brown G. "5G use cases & candidate technologies," 2016. New York (NY): Heavy Reading.

[3] Hedman P. "Description of network slicing concept," 2016. NGMN Alliance. Version 1.0. Available at: https://www.ngmn.org/uploads/media/160113_Network_Slicing_v1_0.pdf.

[4] Marek S. "Dynamic network slicing is the 'secret sauce' behind 5G." SDxCentral. 2016 Aug 30. Available at: https://www.sdxcentral.com/articles/news/dynamic-network-slicing-secret-sauce-behind-5g/2016/08/.

[5] Paul M, Schaller S, Betts M, Hood D, Shirazipour M, Lopez D, Kaippallimalil J. "Applying SDN architecture to 5G slicing," 2016. Palo Alto (CA): Open Networking Foundation. Report No. TR-526. Available at: https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical-reports/Applying_SDN_Architecture_to_5G_Slicing_TR-526.pdf.

[6] Brown G. "Cloud RAN: New network architectures & industry structures," 2016. New York (NY): Heavy Reading.

5G and the Auto

Staff Writer

[Photo credit: Chesky_W/iStock/Thinkstock]

Industry stakeholders envision 5G as a key enabler that allows network connectivity in vehicles to shift in status from an optional accessory to a core feature that supports not only the individual vehicle, but also communication with other vehicles and sensors that inform traffic, parking, and navigation—while also ensuring passenger safety and data security. For consumers, a connected vehicle provides a growing number of features and services that make the driving experience safer, convenient, and less costly. 5G connectivity will enable information from in-car sensors to continuously be passed to the cloud. By sharing information and alerts about micro-level weather, road temperature, surface conditions, and violent breaking ahead, more efficient and consistent traffic flows will be achieved that reduce congestion and emissions. The aggregated and interpreted data will provide more informed driving information, as well as alert and activate onboard safety systems to prevent accidents [1].

motive Industry

Autonomous vehicles

Autonomous vehicles—also referred to as driverless or piloted vehicles—are predicted to hit the market by 2020, but stable 5G infrastructure will play a key role, according to a white paper by ABI Research [2]. Estimates indicate that 5G latency could be as low as one millisecond (ms) over-the-air, and 5 ms end-to-end, enabling the following automotive use cases:

- Broadband multimedia streaming (driverless vehicles as mobile living rooms).
- Cloud services for vehicle lifecycle management, apps, security, and over-the-air updates (cloud-to-vehicle).

- Capturing or uploading huge volumes of sensor data for real-time traffic, weather, parking, and mapping services (vehicle-to-cloud).
- Cooperative mobility: low latency vehicle-to-vehicle and vehicle-to-infrastructure for active safety and autonomy [redundancy for advanced driver assistance systems (ADAS)] [2].

The current ADAS being delivered on 2016 vehicles already facilitate SAE Level 1 (see figure 1) and are beginning to incorporate features that would be considered Level 2. However, Level 4 and 5 capabilities may not be that far away. In October 2015, Robot Taxi, a joint venture between Japanese mobile Internet company DeNA and vehicle technology developer ZMP,

announced that it would offer driverless transportation to about 50 people in an area near Tokyo. Its goal is to commercialize the service by 2020, in time for the Tokyo Olympics [1]. While Robot Taxi is shooting for full automation (Level 5) in time for the games, it is likely that the use of such vehicles will initially be limited to shuttling passengers between Olympic venues. However, this venture is particularly notable because the technology is brand-agnostic and can be retrofitted to any vehicle [4].

Clearly, experimentation with autonomous vehicles is increasing, and 2020 as a date for some form of commercial implementation is certainly feasible from a technology perspective. Probably the best-known self-driving car project belongs to Google, which was started in 2009 and has clocked more than two million miles on public roads to date [5]. However, Google is not alone; Tesla, BMW, Audi, Mercedes, and most recently GM, have all showcased self-driving concept cars and demonstration projects. In mid-October 2016, Tesla announced that all cars currently being produced in their factories would include the hardware needed for full self-driving capability at a level of safety far greater than that of a human driver. However, the company added that the technology first needed to be tested and calibrated via "millions of miles of real-world driving" before the hardware would be activated on consumer vehicles [6]. In addition to traditional car manufacturers, companies such as Uber and Chinese search giant Baidu are also working on autonomous technology and self-driving cars [1].

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring of Driving Environment*	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
Human driver monitors the driving environment						
0	No Automation	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
Automated driving system ("system") monitors the driving environment						
3	Conditional Automation	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	High Automation	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	Full Automation	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

FIGURE 1. SAE Levels of automation. The Society of Automotive Engineers (SAE) has defined levels of automation to clarify what role (if any) drivers have in operating a vehicle while a driving automation system is engaged. These levels are intended to establish a consistent framework that can be used across industries as the dialogue about autonomous vehicles continues. (Figure credit: SAE International J3016 [3].)

The motivations for creating an autonomous vehicle are beyond just technology. It's about reducing emissions through better fuel consumption, as well as addressing the demographic changes of an aging population that increase, rather than decrease, the potential for human error-induced accidents. It's also about leveraging the convergence of the shared economy and urban living, where young and old people no longer feel the need to own a car if there is a cost-effective and convenient alternative, such as Zip Car rentals, or on-demand ride-sharing services such as Uber [1].

Looking forward

Geographical coverage will be a key condition for 5G to have any relevance in the automotive sector. Initial 5G coverage can be supplemented by 4G and Wi-Fi connectivity on phones and other devices while the infrastructure is being built up, but these multimode, multiconnectivity solutions will not suffice for critical automotive use cases relying on the unique capabilities of 5G in terms of latency, reliability, and security [2]. Even once 5G is fully deployed, the adoption of self-driving technology will likely play out differently in the various markets in different regions since the forces shaping it are diverse at both the global and local level [1].

Government and industry cooperation

The continuous progression of ADAS-enabled cars and the gradual adoption of the autonomous vehicle will significantly reduce, and possibly eliminate, the number of crashes. This could, in turn, allow the removal of some regulations that relate to safety considerations, such as crumple zones, bumpers, and airbags. It also means that a review of laws relating to driving age, drunk driving, and speed restriction enforcement may be required, but not until all vehicles are compliant [1].

In September 2016, the US Department of Transportation (DOT) released the Federal Automated Vehicles Policy for highly automated vehicles (HAVs), or those intended to operate at Levels 3 to 5 as defined by SAE. The document—which is currently intended as guidance rather than formal policy—lays out standards for safe design, development, and testing of HAVs before they are commercially sold or operated on public roads. It also proposes guidelines for state governments to ensure a consistent national framework for regulation of motor vehicles with all levels of automated technology [7]. DOT's National Highway Traffic Safety Administration released additional nonbinding guidance in October 2016, outlining best cybersecurity practices for motor vehicle manufacturers and individuals and organizations involved in developing self-driving technology. The guidance aims to make cybersecurity a top priority for the automotive industry and proposes layered solutions to ensure that automated driving systems are designed to take appropriate and safe actions, even when an attack is successful [8].

A challenge yet to be addressed is that, historically, car manufacturers have completely controlled the design and development of vehicles. The advent of computers and software that effectively become the "mind" of the car means that manufacturers could lose control to technology and software companies, and yet still remain liable for any issues or catastrophes related to the car. It also remains to be seen how much and what type of data car manufacturers and network providers would be expected to share to improve overall safety and security of connected vehicles [1].

References

[1] Bell S. "IOT & the automotive industry," 2015. New York (NY): Heavy Reading.

[2] Bonte D, Hodgson J. "5G in automotive," 2016. Oyster Bay (NY): ABI Research.

[3] SAE International. "SAE International technical standard provides terminology for motor vehicle automated driving systems" [Press release]. 2 Oct 2014. Available at: http://www.sae.org/misc/pdfs/automated_driving.pdf.

[4] Collins K. "Driverless taxis could be the bullet train of the Tokyo 2020 Olympics." CNET. 22 Aug 2016. Available at: https://www.cnet.com/news/driverless-taxis-could-be-the-bullet-train-of-the-tokyo-2020-olympics/.

[5] Higgins T. "Google's self-driving car program odometer reaches 2 million miles." The Wall Street Journal. 2016 Oct 5. Available at: http://www.wsj.com/articles/googles-self-driving-car-program-odometer-reaches-2-million-miles-1475683321.

[6] Tesla. "All Tesla cars being produced now have full self-driving hardware" [Blog post]. 2016 Oct 19. Available at: https://www.tesla.com/blog/all-tesla-cars-being-produced-now-have-full-self-driving-hardware/?utm_campaign=GL_Blog_101916.

[7] US Department of Transportation. "Federal automated vehicles policy." September 2016. Available at: https://www.transportation.gov/AV/federal-automated-vehicles-policy-september-2016.

[8] National Highway Traffic Safety Administration. "Cybersecurity best practices for modern vehicles." October 2016. Report No. DOT HS 812 333. Available at: http://www.nhtsa.gov/staticfiles/nvs/pdf/812333_CybersecurityForModernVehicles.pdf.

[Photo credit: moodboard/Thinkstock]

FROM LAB TO MARKET

News from the Technology Transfer Program

CREATING & GROWING ECONOMIC IMPACT

NSA's Technology Transfer Program (TTP) is igniting innovation through collaborations with industry, academia, and other government agencies that introduce NSA technologies to the commercial marketplace. By fulfilling the legislative mandate to share government-patented technologies with industry, NSA's TTP plays a role in strengthening national security by contributing to the nation's economic growth.

Technology Transfer Program
Office of Research & Technology Applications
NSA Research Directorate

tech_transfer@nsa.gov
www.nsa.gov/techtransfer
866-680-4539

2000–2014 REPORT

TOTAL DIRECT SALES
from new products and services

$37 MILLION

$120 MILLION

~**3X** INCREASE IN VALUE
Compared to 2000-2011

TOTAL ECONOMIC OUTPUT NATIONWIDE
including direct, indirect, and induced impact

$118 MILLION

$346 MILLION

0   25   50   75   100   125   150   175   200   225   250   275   300   325   350

JOBS CREATED
or retained

1,330

 = 50 jobs

SOURCE: *National Economic Impacts from DoD License Agreements With U.S. Industry*, TechLink and UC Business Research Division.

THE RESULTS

NSA Technology + Technology Transfer = Economic Growth ⇒ National Security

Article

IoT's Tiny Steps towards 5G: Telco's Perspective

Enida Cero [1], Jasmina Baraković Husić [1,2,*] and Sabina Baraković [3,4]

[1]   BH Telecom, Joint Stock Company, Sarajevo, Franca Lehara 7, 71000 Sarajevo, Bosna i Hercegovina;
    enida.cero@bhtelecom.ba

[2]   Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne bb, 71000 Sarajevo,
    Bosna i Hercegovina

[3]   Faculty of Transport and Communications, University of Sarajevo, Zmaja od Bosne 8, 71000 Sarajevo,
    Bosna i Hercegovina; barakovic.sabina@gmail.com

[4]   Department for Computer Engineering, American University in Bosnia and Herzegovina,
    Trg solidarnosti 10, 71000 Sarajevo, Bosna i Hercegovina

*   Correspondence: Jasmina.BarakovicHusic@bhtelecom.ba; Tel.: +387-61-265-500

Abstract: The numerous and diverse applications of the Internet of Things (IoT) have the potential to change all areas of daily life of individuals, businesses, and society as a whole. The vision of a pervasive IoT spans a wide range of application domains and addresses the enabling technologies needed to meet the performance requirements of various IoT applications. In order to accomplish this vision, this paper aims to provide an analysis of literature in order to propose a new classification of IoT applications, specify and prioritize performance requirements of such IoT application classes, and give an insight into state-of-the-art technologies used to meet these requirements, all from telco's perspective. A deep and comprehensive understanding of the scope and classification of IoT applications is an essential precondition for determining their performance requirements with the overall goal of defining the enabling technologies towards fifth generation (5G) networks, while avoiding over-specification and high costs. Given the fact that this paper presents an overview of current research for the given topic, it also targets the research community and other stakeholders interested in this contemporary and attractive field for the purpose of recognizing research gaps and recommending new research directions.

Keywords: 5G mobile communication; Internet of Things; applications classification; performance requirements; enabling technologies

---

1. Introduction

Telecom operators (telcos) had the most power and influence over business in the information and communication technology (ICT) industry during past decades. This dominance was the result of owning and provisioning communication infrastructures which nowadays have become more of a commodity than a luxury. Telco's revenue streams were mainly based on the provisioning of traditional services, such as voice calls and short message services (SMS). Recent work of regulatory agencies and the appearance of alternative service providers have led telecom operators to form the opinion that alternative service providers have conflicting interests and provide competitive services, thus decreasing telco's revenue from traditional services [1]. However, the latest econometric analysis presented in [2] has shown that telecom operators and alternative service providers have aligned interests and their collaboration could be favorable to both parties. The interests of alternative service providers and telcos are not inevitably conflicting, since the economic growth of the alternative service providers are positively correlated with telco revenues and vice versa.

The development of a digital society has changed the traditional value chain and introduced new issues for telcos as they seek a way how to monetize new digital services [3]. These services

are diversified across various domains of industry, such as agriculture, construction, utilities, transportation, healthcare, finance, etc., and delivered to customers through connected devices building thereby the concept of Internet of Things (IoT).

IoT can be defined as an interconnected network of things/objects that are able to interact with each other and cooperate with other things/objects through wireless and wired connections in order to create new services/applications for the benefit of society. In this way, IoT brings almost limitless benefit's, which have the potential to radically change our daily life by saving time and resources while creating possibilities for growth and innovation [4]. IoT has disruptive potential in almost every application domain, which has created many challenges that need to be faced when implementing IoT-based solutions, such as prevailing over obstacles generated by the fragmentation of IoT, both in terms of application domains and in terms of technologies. Therefore, recognizing all potential IoT applications while having in mind the development of technology and the requirements of individuals, businesses, and society as a whole is quite challenging.

Telcos now have the opportunity to seize a share of the value that is generated by IoT implementation. The size of this share will depend on telco's role in the value chain that ranges from being a traditional provider of communication infrastructure to being an end-to-end solution provider [3]. In order to monetize IoT, telcos will have to address many challenges which can be summarized as follows: (1) strategic challenges relating to decisions on future directions; (2) business challenges relating to successful management, investment, partnerships; and (3) technical challenges relating to changing connectivity and performance requirements. This paper focuses on the technical challenges since connectivity and performance requirements of IoT objects/things cannot be fulfilled using existing cellular networks that limit numerous IoT applications. In order to overcome issues associated with the current cellular networks, new types of technologies are being introduced leading towards the fifth generation (5G) network [5–7].

The architecture of the 5G network has to seamlessly integrate the requirements of diverse IoT applications: from delay-sensitive video applications to ultra-low latency, from high-speed entertainment applications in a vehicle to mobility on demand for connected objects and from best effort applications to reliable and ultra-reliable ones for health and safety [8]. A full understanding of emerging IoT applications and the variability of their performance requirements can serve telecom operators as input for specifying 5G enabling technologies. These technologies should be flexible and scalable to meet the aforementioned requirements. Since IoT applications sometimes demand extreme requirements, the 5G network must simultaneously satisfy all of them, which can lead to over-specification and high cost. In order to avoid this, telecom operators first have to adequately classify IoT applications to facilitate the selection of 5G enabling technologies being capable to efficiently meet their performance requirements.

The 5G classification concept includes three different service classes [9], i.e., (1) extreme mobile broadband (xMBB); (2) massive machine-type communications (mMTC); and (3) ultra-reliable machine-type communications (uMTC). Nevertheless, this classification concept can be considered insufficient to properly select 5G enabling technologies to meet the diverse requirements of IoT applications. Furthermore, it will be challenging to classify the emerging IoT applications given that there is many criteria for their categorization. The definition of the IoT domain in many cases overlaps with the definition of IoT application, which may at the same time belong to another IoT domain. For example, Smart Buildings may either be considered as a standalone IoT application or as an element needed to form a Smart City application domain. IoT applications are usually classified according to spheres of human life [4,10–12] or performance requirements [9,13–15]. However, these classifications are not suitable for telecom operators which have to fulfil the performance requirements of particular IoT applications, since it is difficult to choose the most important one from a class containing a wide range of IoT applications.

Telecom operators can easily specify performance requirements for current communication services since traffic patterns generated by these services are driven by predictable activities, such as

making calls, receiving email, surfing the web, and watching videos. However, traffic patterns generated by emerging IoT applications are driven by less predictable activities. Hence, the idea is to determine a common set of activities based on customer service requests which describe the reasonably foreseeable function/purpose of emerging IoT applications. Then these activities can be used by telcos as a new classification criterion of IoT applications. The assumption is that such a classification can help telecom operators meet performance requirements of future IoT applications and select appropriate 5G enabling technologies. Therefore, the aims of this paper are to classify IoT applications according to relevant activities, specify and prioritize the performance requirements of each IoT application class, and consider enabling technologies used to accomplish specified demands on the radio access part of 5G networks, all according to telco's perspective. In addition, the intention is to provide a literature review which covers 258 references dominantly published in the period between 2012 and 2017. As this paper provides a state-of-the-art review of IoT applications, their performance requirements and 5G enabling technologies, it can also be of use to the research community and other stakeholders interested in this contemporary and attractive field in order to recognize the research gaps and recommend new research directions.

The remainder of the paper is thusly organized. In order to answer research questions, Section 2 describes the methodological approach in regard to the conducted research. Section 3 identifies a common set of activities as a new classification criterion for IoT applications and assigns them the relevant 5G service classes. This facilitates the determining of performance requirements of activity-based IoT application classes, which are summarized and prioritized in Section 4. The enabling technologies used to accomplish these requirements in 5G networks are contemplated in Section 5. Finally, Section 6 discusses the outcomes of the paper by clarifying the contribution of our research in identifying open issues for future work, while Section 7 concludes this paper.

2. Research Approach and Design

The main objectives of this paper can be summarized as follows: (1) to identify activities relevant to IoT customer service requests and use them as a new classification criterion of IoT applications; (2) to specify and prioritize performance requirements of such IoT application classes; (3) to analyze relevant literature in order to provide insight into 5G radio technologies used to fulfil the requirements of activity-based IoT application classes; (4) to recognize research gaps and directions.

The research questions posed in this study are: (1) Which activities can be used as classification criteria for IoT applications? (2) What are the performance requirements of such IoT application classes? (3) Which enabling technologies in the radio access part of the 5G network can be used to meet the requirements of IoT application classes?

Investigation of different criteria for the classification of IoT applications underlines to what extent various classification concepts have been covered by existing literature and provides a basis for the introduction of a new classification criterion called the activity. We thus surveyed and compared different studies contributing to the understanding of the scope and classification of a wide range of IoT applications. This approach has been motivated by the challenge to propose an activity-based classification of IoT applications, the determination and prioritization of performance requirements of such IoT application classes and the review of 5G enabling technologies used to meet them. These findings can be utilized by telecom operators and other interested parties (e.g., the research community, software network function providers, network infrastructure manufactures, etc.) depending on their interest and potential to utilize these findings towards 5G implementation and commercialization.

The methodological approach to research conducted in this paper is illustrated in Figure 1 and includes four phases.

Phase I included searching, identifying, and extracting papers from three categories, i.e., IoT in 5G service classification, IoT in 5G performance requirements and IoT in 5G enabling technologies. The keywords used to search relevant scholar databases are shown in Figure 2. This search has

resulted in the selection of 258 recently published papers, which can be categorized into three groups (i.e., review papers, technical papers, research effectiveness), as shown in Table 1. Review papers summarize the status of knowledge and outline future directions in a given area of research. Technical papers describe the process, progress, or result of research, whereas research effectiveness directly provides an answer to the research question raised for this study. The reference distribution by year of publication is shown in Figure 3a, while the total percentage of references per publication type is presented in Figure 3b.
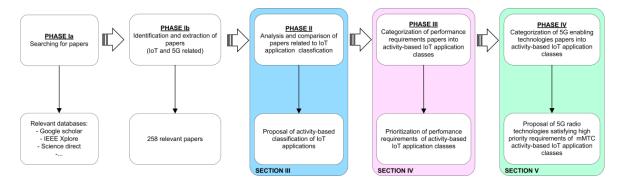
Figure 1. Methodological approach to research.
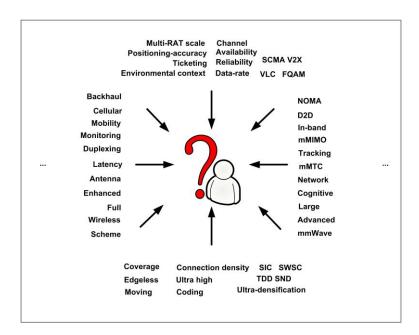
Figure 2. Keywords for searching relevant databases.

Table 1. Reference categorization.

Reference Type	Reference Number
Review paper	[1,2,10,12,14–88]
Technical paper	[8,89–197]
Research effectiveness	[3–7,9,11–13,29,52,198–258]
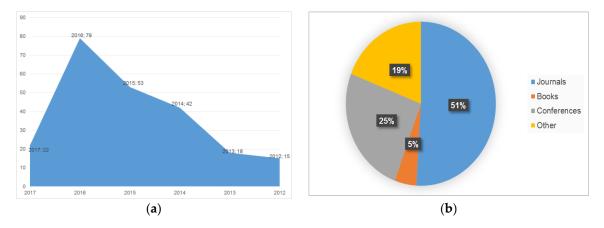
Figure 3. Reference distribution by (**a**) year of publication; (**b**) type of publication.

Phase II included the analysis and comparison of papers related to the category of IoT in 5G service classification. Papers were first sorted according to the existing classification criteria of IoT applications in order to underline their deficiencies from telco's perspective. Since emerging IoT applications generate traffic driven by less predictable activities, specifying their performance requirements and enabling technologies is not an easy task from telco's point of view. Therefore, we have proposed a common set of activities to serve as classification criteria of IoT applications. Four activities were selected based on IoT customer service requests denoting the function/purpose of existing IoT applications. According to the authors' knowledge, existing IoT applications can be grouped around following activities: ticketing, tracking, monitoring, and managing/controlling. Such a classification makes it easier for telecom operators to associate particular IoT applications to the relevant 5G service class. Finally, Phase II has resulted in the proposal of an activity-based classification of IoT applications which was associated with the 5G service classification.

Phase III included the analysis of papers related to the category of IoT in 5G performance requirements. Papers were grouped according to eight key performance indicators, i.e., data rate, mobility, latency, connection density, reliability, positioning accuracy, coverage, and energy efficiency. These performance requirements were prioritized indicating high, medium, and low importance of each requirement for specific activity-based IoT application classes proposed in Phase II. Finally, Phase III resulted in the identification of highly important performance requirements for each activity-based IoT application class which can be utilized by telcos to identify 5G enabling technologies used to meet them.

Phase IV provided a review of papers from the category of IoT in 5G enabling technologies. Papers were first sorted according to eight technological groups used in the radio access part of 5G networks, i.e., wide and flexible bandwidth technology, advanced modulation and coding, duplexing, multiple access and waveform, advanced interface management, access architecture related radio technologies, energy related technologies, and other technologies. These technologies have been discussed in terms of their possibility to meet the high priority performance requirements of mMTC activity-based IoT application classes, since this area is more mature from telco's perspective. The results of Phase III were reflected in a proposal of technologies in the radio access part of 5G networks that can be used to meet the high priority performance requirements of the mMTC activity-based IoT class, as well as the identification of research gaps and directions for future work.

3. IoT in 5G Service Classification

As mentioned in the Introduction, various challenges awaiting telcos can be identified on the basis of the role that they will play in the IoT value chain. Regardless of this role, telcos will face many technical challenges, such as the necessity for global deployment, the need for rapid infrastructure scaling, unpredictable IoT application behavior, etc. Even now, existing IoT applications accelerate

the growth rate of traffic driven by less predictable activities which require new strategies towards 5G networks. These networks will be the key enabler for IoT applications by providing a unified infrastructure capable of meeting the high variability of their performance requirements [8,16]. Various stakeholders have recently described what 5G networks may be and grouped the major types of 5G services into three different classes: (1) xMBB—Requiring extremely high data rates and low-latency communication; (2) mMTC—Requiring scalable connectivity for an extremely large number of devices; and (3) uMTC—Requiring ultra-reliable low-latency and resilient communication [198]. This 5G service classification can be used by telcos to identify performance requirements of various IoT applications, and, as such, was used in this research study. However, emerging IoT applications will have extreme requirements, so an analysis based solely on 5G service classifications may not be sufficient. Therefore, telcos need a more precise classification of IoT applications which is an important precondition for meeting their diverse requirements.

IoT covers human-to-human (H2H), human-to-machine (H2M) and machine-to-machine (M2M) communication, which will be the main driving force towards 5G networks. In addition, the terms M2M communication (M2MC) and machine type communication (MTC) are used interchangeably as in [17–19,89–93,199,200]. Although IoT is a broader concept which evolves from M2M, this paper assumes that IoT and M2M are synonyms as in [94].

3.1. Existing Classifications of IoT Applications

In order to meet the requirements of a wide range of IoT applications, they have to be classified in an appropriate manner. The existing approaches to the classification of IoT applications are summarized in Table 2. IoT serves different user categories, including individuals, businesses, and society as a whole, and may span through a broad range of application domains [4,10–12], such as transportation and logistics, healthcare, smart environment, personal and social, futuristic applications, food sustainability, smart living, smart manufacturing, smart energy, smart city, etc. These application domains are created to be human-centric, which means that they cover different domains of human life. IoT applications belonging to these domains have diverse requirements for 5G networks, and should not be treated equally. Therefore, the acceptance of these domains as a classification criterion raises the issue of assigning IoT application to a particular domain. Moreover, the emergence of new applications of IoT may require defining new application domains, which makes this classification inappropriate from telco's perspective.

The M2M applications may be classified by the mobility and the amount of dispersion that needs to be supported into four categories [13]: (1) fixed and concentrated; (2) fixed and dispersed; (3) mobile and concentrated; and (4) mobile and dispersed. However, the mobility and amount of dispersion present very rough classification criteria and cannot meet the precise network requirements of individual IoT applications.

Additionally, M2M applications can be grouped according to delay tolerance into four categories [14]: (1) elastic (delay tolerant); (2) hard real-time (delay constraint); (3) delay-adaptive (delay sensitive but tolerant); and (4) rate-adaptive application (adjust their transmission rates according to available radio resource). However, the main drawback of this classification is the lack of consideration of other IoT application requirements except delay tolerance, although its importance as a classification criterion has been recognized.

According to data reporting mode, the M2M applications can be classified into five categories [91]: (1) time-driven; (2) query-driven; (3) event-driven; (4) continuous-based; and (5) hybrid-driven. This classification is specific for the former IoT concept, which is narrower in nature than the definition adopted in this study.

Another study considered the reliability, availability, and end-to-end latency in order to classify IoT applications into two groups [15]: (1) monitoring-based and mission-critical; (2) monitoring-based and non-mission critical; (3) control-oriented and mission-critical; and (4) control-oriented and non-mission critical. Monitor-based IoT applications periodically collect sensor data from smart

objects and transmit them. The majority of monitoring-based IoT applications are not mission-critical. Control-oriented IoT applications use sensor data to control actuators in real-time, and rely on mission critical communication. This classification is based on multiple criteria and represents the precursor of the Mobile and wireless communications Enablers for Twenty-twenty Information Society (METIS) 5G service classification.

Table 2. Internet of Things (IoT) application classification—summary.

Criteria for Classification	IoT Classes	References
Domains	(1) Transportation and logistics (2) Healthcare (3) Environment (4) Personal and social (5) Futuristic applications (6) Food/water monitoring (7) Living (8) Manufacturing (9) Energy (10) Building (11) Industry (12) City (13) Security and safety (14) Communication (15) e-society (16) Vehicular (17) Sport and leisure	[4,10–12]
Mobility and amount of dispersion	(1) Fixed and concentrated (2) Fixed and dispersed (3) Mobile and concentrated (4) Mobile and dispersed	[13]
Delay tolerance	(1) Elastic (2) Hard real time (3) Delay-adaptive (4) Rate-adaptive	[14]
Data reporting mode	(1) Time-driven (2) Query-driven (3) Event-driven (4) Continuous-based (5) Hybrid-driven	[198]
Reliability, availability, and end-to-end latency	(1) Monitoring-based and mission critical (2) Monitoring-based and non-mission critical (3) Control-oriented and mission critical (4) Control-oriented and non-mission critical	[15]
Characteristics and requirements	mMTC and uMTC	[12]

Legend: IoT (Internet of Things), mMTC (massive Machine Type Communication), uMTC (ultra-reliable Machine Type Communication).

Although it spans through a wide range of different applications, MTC can be divided in two main categories, i.e., massive and ultra-reliable MTC, which depend on their characteristics and requirements [89]. As mentioned above, this categorization is a part of 5G service classification which was used in this study as a basis to identify performance requirements of IoT applications and will be described in more detail in Section 4. Massive MTC (mMTC) typically involves a very large number of devices (tens of billions [9]), such as sensors, actuators, and similar devices [12], different in complexity and cost [9], and with varying quality of service (QoS) requirements. These devices should be of very low cost with very low energy consumption, enabling very long battery life [12]. At the same time, the amount of data generated by each device is normally very small, and very low latency is not a critical requirement [12]. Ultra-reliable MTC (uMTC) requires very high reliability and availability, and very low latency [9,12]. Low device cost and energy consumption are not as critical as they are for mMTC applications [12], and the number of devices and required data rates are relatively low [198].

The vision of a pervasive IoT requires the integration of various domains into a single and unified domain. It addresses the enabling technologies needed for these domains while taking into account the elements that form the third dimension like security, privacy, trust, and safety [11]. The current classification criteria and IoT application classes do not clearly differentiate IoT domains from IoT applications. IoT domains are usually viewed as a specific area of IoT applications that support a huge variety of use cases across industries impacting businesses and customers. Therefore, IoT domains overlap with IoT applications within existing IoT classifications. Moreover, IoT classification based on certain QoS parameters is difficult to apply from telco's perspective, because it is challenging to choose the most important parameter for the wide range of IoT applications. However, these problems need to be solved for upcoming 5G networks.

3.2. Activity-Based Classification of IoT Applications

The aforementioned issues motivated us to propose a new approach to the classification of IoT applications. This approach is based on "**the activity**", which primarily characterizes specific IoT application. In this sense, the activities are defined as new classification criteria, which denote a main function/purpose of specific IoT application observed from telco's point of view. Telecom operators observe the technical challenges of IoT through systems, tools, devices, and platforms because their availability and integration complexity determines the opportunity to capture a share of the value that is generated by IoT implementation. Based on the literature review, it was found that the most commonly mentioned terms in this sense were related to: (1) ticketing system [201]; (2) monitoring: devices [199,201–204]/services [199,201,202,205,206]/tools [207]/systems [11,12,199,201,205,206,208–210]/ data [95,201]/framework [12,201]/solutions [20,205]/networks [211]/process [7]/activity [203,204,207]; (3) tracking technologies [4]/devices [204]/systems [204]/applications [11,210]/system [201]/services [202,205]; (4) managing/controlling applications [15,210,212]/operation [211]/services [4,208]/tools [208,211]/devices [204]/system [95,198,204,210]/concepts [211]/solutions [95]/platforms [95,199]. This has led us to identify four activities, i.e., ticketing, monitoring, tracking, and managing/controlling, as new classification criteria of IoT applications. According to the authors' knowledge, it was found that these activities can cover reasonably foreseeable functions/purposes of IoT in existing application domains. Using activities as classification criteria, telcos can specify IoT application performance requirements more easily, which are in that case dictated by the specific activity, not by the IoT application domain, and determine enabling technologies in the radio access part of 5G networks.

In addition, our classification approach allows better service differentiation and service delivery closer to customer expectations. According to BH Telecom's experience, customers typically come up with the following service requests when it comes to IoT: (1) they need to track their products and determine products' distribution across different regions based on their own data analysis; (2) they want to monitor their products since they do not have a department for supervision and analysis; (3) they need information panels about their outlets and working time to be downloaded by a scanning tab-ticket; (4) they want to manage/control their products according to market needs. This way of expressing customer needs has inspired us to propose the activity-based classification of IoT applications as it allows telcos to define performance requirements more precisely, and thereby improve the customer experience.

The activity-based classification of IoT applications is presented in Table 3. Each activity is associated with the application domain where that activity may be applied. The requirements of identified activities and associated application domains are then mapped to 5G service classes in order to be further able to identify their performance requirements and enabling technologies necessary to meet them. For example, according to Table 3, the managing/controlling activity can be realized in several domains, such as healthcare, food, energy, transportation and logistics. For each domain, we have identified an application example, such as remote surgery [213] in the healthcare domain, food processing facilities [96] in the food domain, energy distribution [12] in the energy domain and traffic/driving [12,213] in the transportation and logistics domain. According to diverse

performance requirements of these IoT applications [214], managing/controlling activity in healthcare, food, transportation, and logistic domains can be associated with the uMTC service class, while the energy domain can be mapped to the mMTC service class. A more detailed description of performance requirements for each activity-based class of IoT applications is provided in Section 4, while in Section 5 it is discussed how these requirements can be accomplished in the radio access part of 5G networks.

Table 3. Activity-based classification of IoT applications.

Activity	Domain Examples	5G Service Classification	Application Examples
Ticketing	Smart Transportation and Logistics	mMTC	POS Terminal [4]
Monitoring	Smart Healthcare	uMTC	Health condition [97]
	Smart Buildings	uMTC	Structures (buildings, tunnels, etc.) [12,25,213]
	Smart Buildings/Smart City	mMTC	Parking spaces
	Smart Buildings/Smart Environment	mMTC	Home video [12]
	Smart food/water monitoring	mMTC	Food growth condition [10]
Tracking	Smart Healthcare/Sport and Leisure	mMTC	Medical assets, wearables [200,213]
	Smart Transportation and Logistics	mMTC	Transport fleet [200,213]
	Smart Industry/Social Networking	mMTC	Shipping of products
	Smart Healthcare	mMTC	People in science museum
Managing/ controlling	Smart Healthcare	uMTC	Remote surgery [213]
	Smart food/water monitoring	uMTC	Food processing facilities [96]
	Smart Transportation and Logistics	uMTC	Traffic, driving [12,213]
	Smart Energy	mMTC	Energy distribution [12,98]

Legend: 5G (Fifth Generation), POS (Point Of Sale), mMTC (massive Machine Type Communication), uMTC (ultra-reliable Machine Type Communication).

As such, the activity-based classification of IoT applications can be used for the creation of new business models, which represent the stakeholder's plan to generate revenue and make a profit from operations, and thereby include many components and functions of the business [21]. However, there is no common opinion which components constitute a business model. The business model architecture can be illustrated by four dimensions [22]: (1) who, identifying the definition of the target customer as one central dimension in designing a new business model; (2) what, describing what is offered to the target customer; (3) how, referring to the construction and distribution of the value proposition; (4) value, explaining why the business model is financially viable. Answering these four questions allows the creation of IoT business models. According to this business model definition, the activity-based classification of IoT applications affects the who and the how dimensions in the following manner. In terms of the who dimension, it directly allows customer segmentation according to considered activities as a way to express their requirements (e.g., customers that require tracking or monitoring of their products). Being aware of the current customer requirements, the proposed classification of IoT applications around four activities (i.e., ticketing, monitoring, tracking, and managing/controlling) can be considered complete. At the same time, the activity-based classification of IoT applications is flexible since additional activities as classification criterion can be concerned with emerging customer requirements. On the other hand, the activity-based classification of IoT applications indirectly impacts the how dimension of new business models which among others includes relevant resources and capabilities in the focal stakeholder's internal value chain. This indicates that the proposed activity-based classification of IoT applications can be easily applied to IoT business models.

Various categories of IoT business models can be identified according to eight IoT architectural layers [23], i.e., collaboration and processes layer, application layer, service layer, abstraction layer, storage layer, processing layer, network communication layer, physical layer. Along with IoT architectural layer, the IoT business model needs also to address the IoT value proposition [24] and

IoT stakeholders that can participate in more than one layer. Based on their role in the IoT ecosystem, telecom operators usually take a part in the network communication layer. In this regard, they follow four evolutionary business models [23]: (1) selling connectivity services only; (2) selling third-party products; (3) selling internal products; (4) providing a broad menu of IoT products. Each of these IoT business models includes traditional telcos strength, i.e., connectivity which determines the how dimension of the IoT business model. In this sense, our activity-based classification of IoT applications indirectly allows the construction and distribution of performance as IoT value proposition using enabling technologies as IoT resources in telco's internal value chain.

3.3. Summary of IoT in 5G Service Classification

The previous discussion has shown that the classification of IoT applications is a complex task due to their numerosity and diversity. The existing classifications of IoT applications pose some drawbacks that can be summarized as follows. The domain-based classification does not allow clear differentiation between IoT domains and IoT applications due to either an imprecise classification criterion or the diversity and unpredictability of IoT applications. On the other hand, QoS-based classifications need to identify the common and most important performance metric for a broad range of IoT applications. Therefore, on the basis of the literature review and IoT customer service requests, we have proposed a new approach to the classification of IoT applications. It is based on the activity as new classification criterion, which denotes a main function/purpose of specific IoT application observed from telco's point of view. According to the authors' knowledge, four activities, i.e., ticketing, monitoring, tracking, and managing/controlling, have been identified to cover reasonably foreseeable functions/purposes of existing IoT applications. The resulting activity-based IoT application classes have been associated with 5G service classes in order to determine and prioritize their performance requirements as described in next section. Finally, the proposed classification of IoT applications was discussed in terms of its completeness, flexibility, and applicability to new business models.

4. IoT in 5G Performance Requirements

This section provides an insight into performance requirements of activity-based classes of IoT applications proposed in Section 3. The analysis is based on eight key performance indicators identified in [5,207] as shown in Table 4: data rate, mobility, latency, connection density, reliability, positioning accuracy, coverage, and energy efficiency. These performance indicators are usually well described for specific IoT applications. However, one of the main challenges of 5G is to support a variety of performance requirements for numerous IoT applications in a flexible, reliable, and cost-effective way [15]. Hence, there is a need for a comprehensive understanding of these requirements for activity-based IoT application classes. The 5G service classification defined the performance requirements for mMTC and uMTC [9]. We have assigned these requirements to the activity-based IoT application classes introduced in Section 3 for the purpose of proposing priorities of each requirement for a specific class. Three levels of priorities (high, medium, low) are associated with the performance requirements of activity-based classes of IoT applications as shown in Table 4. The prioritization of performance requirements is inspired by analysis of related work undertaken in [213]. Some activity-based classes of IoT applications may demand optimization of multiple performance requirements. Table 4 illustrates the main differences between activity-based IoT application classes, and therefore, the need for a 5G network that enables support of optimal configurations for a variety of, sometimes opposite, requirements. For example, the mMTC tracking-based IoT application class requires support for high mobility, high positioning accuracy, and high connection density, while the mMTC monitoring-based IoT application class also requires high connection density but low mobility and low positioning accuracy. A more detailed description of considered performance requirements is contained in the following subsections.

Table 4. Activity-based classes of IoT applications—performance requirements.

5G Service Classification	Activity	User Experienced Data Rate [Gbps] Outdoor: 0.1 [5] Indoor: 1 [5]	Mobility [km/h] Required: 500 [5]	Latency [ms] Control Plane: 50 [5] User Plane: 1 [5]	Connection Density [Connections/km^2] Required: 10^6 [5]	Reliability [%] Required: 99.999 [207]	Positioning Accuracy Required: A Few cm [5]	Coverage/ Availability [%] Required: 99.999 [207]	Energy Efficiency [bits/J]
mMTC	Ticketing	L to M	L	L	H	M	H	H	H
	Tracking	M	H	L	H	M	H	H	M to H
	Monitoring	M to H	L	L	H	L to M	L	H	M
	Managing/controlling	L	L	L	H	M	L	M to H	M
uMTC	Monitoring	L	L to M	H	H	H	M to H	H	M
	Managing/controlling	L	M to H	H	L	H	H	M to H	M

Legend: 5G (Fifth Generation), mMTC (massive Machine Type Communication), uMTC (ultra-reliable Machine Type Communication), L (Low), M (Medium), H (High).

4.1. Data Rate

Data rate is the most important evaluation factor for generations of wireless communication networks [6]. It is contemplated in two ways: (1) peak data rate—defined as the maximum achievable data rate by the user; and (2) minimum guaranteed user data rate—defined as the minimum experience data rate by the user [16]. New mobile technologies are primarily driven by users' needs for higher data rates, as discussed in [8,16,99,215,216]. The expected values in 5G networks are 10 Gbps for minimum peak data rate and 100 Mbps as minimum guaranteed user data rate [5]. High data rate requirements are mainly posed by xMBB related use cases [9] like hotspots [214] or dense urban areas where in 95% of locations and time experience data rate by user should be 300 Mbps in downlink and 60 Mbps in uplink [100]. High data rate is also important in some activity-based classes of IoT applications, as shown in Table 4. Medium (e.g., monitoring of parking spaces) to high (e.g., monitoring of home video) data rate is needed in cases of mMTC monitoring activities. High data rate can be achieved by using a millimeter wave (mmWave) spectrum to share multi-gigabit data in the surrounding environment (i.e., mMTC monitoring with high data rate requirement) and to recognize an object via cloud in real time to find the optimal driving strategy instantaneously for IoT—autonomous vehicles application [101]. Data rates generated during transmissions of tracked medical assets, transport fleets, and ticked point of sale (POS) terminals is low (e.g., medium importance for ticketing and tracking activities in mMTC). In addition, low data rates are needed in uMTC monitoring and controlling (e.g., monitoring of health condition, control of driving).

4.2. Mobility

Mobility is defined as relative velocity between the receiver and the transmitter [16]. The applications of IoT pose very diverse requirements for mobility in 5G networks, which range from static to high mobile, even up to 500 km/h [16,102,207]. Use cases in which except ultra-high mobility, ultra-high traffic volume density, and ultra-high connection density are needed may be quite challenging for 5G networks [215], like V2X communication [5]. High mobility is a very important requirement for mMTC tracking activity (e.g., tracking of assets in high speed trains [214]). Moreover, the support for high mobility is needed in uMTC management activities, if the monitor/control object is moving (e.g., high speed trains [207]). Low mobility is needed for ticketing, monitoring, and managing/controlling activities in the mMTC class, as shown in Table 4. This is the reason why 5G networks should not assume mobility for all devices and services but rather provide mobility on demand [8].

4.3. Latency

Latency requirements are usually expressed in terms of end-to-end (E2E) latency perceived by the end user [8]. 5G networks should enable "zero latency" [103] represented by the millisecond level of E2E latency [16,26,27,213] through significant enhancements and new technology in architecture aspects [6,28,214], such as device-to-device (D2D) communication [5]. The IoT applications associated with uMTC monitoring and managing/controlling activities require low latency, as they tend to be real time. Required latency levels depend on the particular IoT application [8,16,28], being the narrowest in uMTC managing/controlling activity with a value that should not exceed 1 ms [8]. For example, the tolerable delay for use case mobile health [215] and remote surgery application is in order of sub-milliseconds [213]. For V2X communications, latencies should be ultra-low for some warning signals [8].

4.4. Connection Density

Connection density is defined as the number of connected devices per unit area [16]. It can usually be expressed in terms of an extremely high number of simultaneous active connections, such as 1 million connections per square meter [207,215], or 10 to 100 times higher number of connected

devices [29]. The performance values of connection and traffic density for various 5G services are listed in [8,212]. This performance requirement is highly correlated with identified activity-based classes of IoT applications, since it is the main characteristic of mMTC. High device density brings the need for rapid internet protocol version 6 (IPv6) deployment, and high quality security algorithms and techniques, which lead to new system design and implementation, as described in [104].

4.5. Reliability

Reliability is the maximum tolerable packet loss rate at the application layer [105]. 5G must bring a reliability of 99,999% [27,30,31,212], or higher for specific use cases [212], (e.g., tele-protection in a smart grid network [207] or driverless cars [106]). Reliability is the main characteristic of uMTC monitoring and managing/controlling activities. Reliability will be a particularly challenging task in high-speed trains because of speed, load, and cell distance [16,107].

4.6. Position Accuracy

Position accuracy is the maximum positioning error tolerated by the application [105]. 5G should ensure accurate positioning of the device outdoors [108] with accuracy from 10 m to <1 m on 80% of occasions and better than 1 m in indoor deployment [212]. Accuracy positioning is very important in uMTC monitoring-based activities (e.g., monitoring remote cameras), and uMTC managing/controlling-based activities (e.g., driving) [213]. Moreover, the mMTC tracking-and-ticketing-based activities pose high performance requirements in terms of position accuracy.

4.7. Coverage

Coverage requirements in 5G should provide connectivity anytime and anywhere with a minimum user experience data rate of 1 Gbps [32]. However, the perception of 100% coverage is rather a technical decision than a business one [33], which could be extended through ultra-cell deployment [31] and vehicle-to-infrastructure (V2I) communications [30]. Almost every activity-based IoT application class requires very high levels of coverage (99,999% availability) [207]. Total coverage will enable new unmanned aerial vehicles (UAV) to use single network connection, instead of connection steering mechanism, as one described in [109].

4.8. Energy Efficiency

Energy efficiency is defined as the number of bits that can be transmitted per joule of energy [216]. Compared with current wireless technologies, the energy efficiency (measured in b/J) of the 5G network may need to be improved by a factor of 1000 [4,7,34,110,217]. High energy efficiency is important in case of ticketing- and in some cases of tracking-based activities (e.g., smart industry [205] or implantable medical devices [35]). Other activities, such as both mMTC and uMTC monitoring- and managing/controlling-based activities (e.g., health condition [205]) and some tracking-based activities (e.g., sports wearables [205]) require medium energy efficiency. Energy efficiency is very important design objective for the reduction of operating costs of telecom operators, as well as for minimizing the environmental impact of the wireless domain [217]. On a higher layer of the network protocol stack, adaptive base station switch on/off algorithms use renewable energy sources to save energy [218] along with an energy scheduler, as in the with heating, ventilation and air conditioning (HVAC) [111]. However, at the physical layer, adaptively switching off unused carriers is a key strategy that can be used to save energy from the radio-frequency (RF) transceiver chain of base stations [218].

4.9. Spectrum Efficiency

Spectrum efficiency is defined as the data throughput per unit of spectrum resource per cell or per unit area (bps/Hz/cellar bps/Hz/km^2) [216]. In order to achieve network sustainability, required for 5G networks [216], spectrum efficiency needs to be improved 3–5 times [20,32,216]. Minimum

peak spectrum efficiency is 30 bps/Hz for downlink and 15 bps/Hz for uplink [32]. This is mostly important for the xMBB services class [212,214]. Since this performance requirement is not relevant for activity-based IoT application classes which are associated with the mMTC and uMTC service class, it has not been further considered herein.

4.10. Summary of IoT in 5G Performance Requirements

Activity-based classes of IoT applications are associated with performance requirements of 5G service classes, i.e., mMTC and uMTC. On the basis of the literature review, each activity-based class (ticketing-, tracking-, monitoring-, and managing/controlling-based class) is associated with mMTC service class, while the monitoring- and managing/controlling-based classes are the only ones linked to the uMTC service class. According to the authors' best knowledge, there is no available literature concerning the uMTC ticketing- and tracking-based classes of IoT applications or such IoT customer service requests.

Activity-based IoT application classes pose many performance requirements, which have been discussed in terms of data rate, mobility (speed), latency, connection density, reliability, positioning accuracy, coverage, and energy efficiency. These performance requirements have been assigned three-level priorities (high, medium, low) for the purpose of facilitating identification of the enabling technologies used to fulfill them. Based on the literature review, it was found that each activity-based IoT application class poses high requirements in terms of connection density except the uMTC managing/controlling-based IoT application class. Moreover, ticketing- and tracking-based IoT application classes have high requirements in terms of positioning accuracy, coverage, and energy efficiency. Additionally, positioning accuracy is a highly important performance requirement for the uMTC managing/controlling-based IoT application class, whereas coverage is important for both uMTC and mMTC monitoring-based IoT application classes. Finally, latency and reliability represent highly important performance requirements for uMTC monitoring- and managing/controlling-based IoT application classes. A deep and comprehensive understanding of performance requirements of each activity-based IoT application class may facilitate the selection of 5G enabling technologies needed to meet them as described in the following section.

5. IoT in 5G Enabling Technologies

Activity-based classification of IoT applications proposed in Section 3 allows telecom operators to identify performance requirements of each class relying on 5G service classification, as discussed in Section 4. However, these performance requirements cannot be satisfied for many IoT applications with current cellular (2nd, 3rd, and 4th generation) network technologies, since they limit their potential due to many issues [3], i.e., protocol implementation complexity, poor coverage in non-urban environments, high cost of networking equipment and data transmission. In order to overcome these issues, many dedicated communication technologies are being installed.

From the very beginning of IoT, many proprietary technologies, such as radio-frequency identification (RFID), wireless highway addressable remote transducer (WirelessHART) or Z-Wave, have first appeared, and then, more generic ones, such as Bluetooth, IEEE 802.15.4, IPv6 over low-power wireless personal area networks (6LoWPAN). However, none of these technologies have become a market leader mainly because of technology shortcomings and business model uncertainty [219]. Hence, new solutions, such as low-power wireless fidelity (LP Wi-Fi), low-power wide area (LPWA) or several improvements for cellular M2M communications have become serious candidates for IoT implementation. LP Wi-Fi is an IEEE 802.11ah standard designed to extend the application area of Wi-Fi networks in order to meet IoT requirements (i.e., large number of devices, large coverage range, energy constrains). First performance studies indicate that this standard will support a broad range of M2M scenarios with a required QoS level, and enable scalable and cost-effective solutions. In addition, LPWA networks have been deployed for some time in the form of many different proprietary solutions (e.g., Amber Wireless, Coronis, Huawei's CIoT, LoRa, M2M Spectrum Networks, Sigfox, Weightless,

etc.), but only the LoRa Alliance, Sigfox and Weightless are involved in LPWA standardization activities. Despite some drawbacks which are mainly related to use of an unlicensed spectrum, LPWA networks are expected to become a key enabler for IoT deployment in early market rollouts and for limited IoT applications [219].

The appearance of these IoT communication technologies was considered as competition to current cellular networks from the telco's point of view from the start. But, in the meanwhile, telcos have realized that the aforementioned communication technologies can be utilized to meet the changing connectivity and performance requirements of IoT applications. Therefore, many dedicated communication technologies have already been deployed in various IoT applications. For example, Orange, Swisscom and South Korea (SK) Telecom have built nationwide networks based on LoRa [3], whereas Deutsche Telekom (DT), Vodafone and all three Chinese operators have completed the rollout of narrowband IoT (NB-IoT), as another LPWA standard utilizing existing long term evolution (LTE) networks [36]. In this context, standardization and interoperability becomes critical because there is a need to consider a broad range of connectivity solutions as presented in Figure 4 [37]. One of those solutions includes satellite technologies which integrated with 5G radio technologies form the 6th generation (6G) standard for providing global coverage [38].

Short-range Wireless Technologies		Data Rate	Mobility	Latency	Connection Density	Reliability	Positioning Accuracy	Coverage	Energy Efficiency
1 Bluetooth — SIG 2 RFID — ISO, IEC, ASTM, DASH7, EPCglobal 3 Zigbee — Zigbee Alliance 4 HART — HART Communication Foundation 5 ISA100.11a — ISA 6 Thread — Foundation for the connected home		X	X					X	
802.11 1 Wifi a-n — IEEE 2 Low Rate WiFi ah — IEEE		X	X					X	
Low Power Wide Area 1 LoRa — LoRa Alliance 2 SigFox — Sigfox Foundation 3 RPMA — RPMA Networks 4 UNB — Telensa 5 Weightless — Weightless SIG 6 Qowisio — Qowisio		X	X	X		X		X	
LTE and NB-IoT 1 LTE Cat-4, Cat-1, Cat-0, Cat-M1 — 3GPP 2 NB-IoT — 3GPP		X						X	
5G IoT		X	X	X	X	X	X	X	X

Figure 4. IoT performance requirements and enabling technologies.

In this regard, standardization activities of 5G and beyond are being undertaken by several standards bodies, such as the Institute of Electrical and Electronics Engineers (IEEE), the 3rd Generation Partnership Project (3GPP), the Internet Engineering Task Force (IETF), ITU Radiocommunication Sector (ITU-R), and ITU Telecommunication Standardization Sector (ITU-T). An overview on these standards bodies and their effort to develop communication standards for 5G and beyond are provided in [112]. Due to the envisioned 5G applicability, these standards have defined an air interface that can significantly improve the performance, and network architectures that allow the deployment and coexistence of various 5G technologies. A detailed discussion of these technologies regarding both technological and standardization aspects indicates that 5G can be perceived as the main driving force for enabling the vision of a truly global IoT [219].

In this sense, the technologies described in this section will allow 5G networks to form a unified communication infrastructure for the realization of a wide range of IoT applications. These technologies

will bring performance enhancement which will trigger a complete revolution in almost all spheres of human life creating a new "everything connected" era. Future 5G networks in such an era will face a serious problem regarding a huge number of different service types [220]. To meet the diverse requirements for a huge number of IoT applications, 5G introduces the concept of network slicing to offer programmable network instances [220]. Using network slicing, multiple independent and dedicated network instances can be created within the same infrastructure to run services that have completely different requirements for latency, reliability, throughput, and mobility [220]. This technology primarily targets a portion of the 5G core network, but also indicates that 5G radio access network (RAN) functionalities can be different for each network slice [221]. To deploy network slices, network functions need to be enabled on an on-demand basis, which has driven the use of virtualization and softwarization in 5G core network [221]. Network function virtualization (NFV) is a paradigm that enables that 5G network functions to run in a virtual environment instead of dedicated hardware [221]. The software define networking (SDN) paradigm facilitates isolation of network slices avoiding the traffic of one slice affecting the performance of another slice [221]. These two paradigms make the network much more dynamic, agile, on-demand, and flexible [222,223]. Since the aforementioned technologies are used to deal with a variety of 5G applications, they will not be further discussed. This paper will focus on technologies used to meet performance requirements for specific applications (inter slice performance) which are implemented in the radio access part of 5G networks.

These technologies are briefly described and discussed in terms of their advantages, disadvantages, and research gaps. Given the fact that the realization of particular technology affects multiple performance requirements, tiny steps towards its implementation will allow telcos to deploy 5G networks, and thereby provide a performance improvement in terms of more capacity, lower latency, more mobility, more position accuracy, increased reliability, and availability [6,214,224]. In other words, 5G networks will accommodate many more users and devices while delivering more data [113,114] to each user requiring high data rates [108] in a more energy-efficient way [115,116].

Table 5 summarizes considered radio technologies used to implement future 5G networks being capable of achieving performance requirements identified in Section 4. On the basis of the literature review, one may conclude that the 5G radio access network is crowded with multiple technologies, and there seems to be a duplication of technologies, all eager to grab telco's attention and convince them to buy into the particular choice. As Table 5 shows, multiple technologies and techniques can be used to meet each performance requirement. Since the mMTC activity-based IoT classes are more mature, we have primarily focused on technologies used to fulfil their high performance requirements (i.e., connection density, positioning, and coverage). We have selected 5G enabling technologies considered from telco's perspective as the most representative and promising candidates to meet these requirements. The implementation of these technologies will also significantly affect the fulfillment of uMTC service class requirements, which will lead to complete automation in all spheres of human life. Therefore, the following subsections will discuss these technologies in more detail with the final goal of identifying research gaps and providing recommendations for future work, which are summarized in Table 6.

Table 5. IoT in 5G enabling radio technologies.

Enabling Technologies		IoT in 5G Service Requirements								
		Data Rate	Mobility	Latency	Connection Density	Reliability	Positioning	Coverage	Energy Efficiency	Spectrum Efficiency
Wide and flexible bandwidth technology	mmWave Band Communication and large-scale antenna	[7,202,225–227]		[202]			[233]	[5]	[202]	[7,29,214]
	Heterogeneous Multi-RAT Integration	[5,217,228–232]						[5]		
	Cognitive Radio and Spectrum Sharing							[5]		
Advanced modulation and coding	Advanced Modulation	[52,234–236]	[237,238]	[239]		[237,239]	[237]			
	Advanced Channel Coding	[235]		[235]						[216]
Duplexing	In-band FD			[124]				[5]		[7,29,214]
	Dynamic TDD	[240]		[240]						[9]
Multiple access and waveform	Multiple access			[216]					[218]	[7,9,29,216,256,257]
	New waveform				[5]					[7,9,29,216,256,257]
Advanced interface management	SND and SWSC	[5]								
Access architecture related radio technologies	Advanced small cell	[7,202]					[5]	[5,251]	[7,202,217]	[202,257]
	MN	[202]	[202]					[5]	[202]	
	Enhanced wireless backhaul							[5]		
	D2D	[202,248]	[7,241,244]					[245–247]	[7,202,215]	[7,9,29,202,252]
Energy related technologies	Energy harvesting						[100,253]			
	UAV	[254]				[254]		[183–185]		
Other important technologies	mMIMO	[7,202,249]		[13,202]		[249]		[200]	[217]	[7,9,214,216,257]
	VLC	[202,250]		[202]					[202,258]	
	SIC		[202]	[202,236]						
	V2X			[255]				[255]		

Legend: IoT (Internet of Things), 5G (Fifth Generation), mmWave (Milimeter-wave), RAT (Radio Access Technology), FQAM (Frequency Quadrature Amplitude Modulation), FD (Full Duplexing), TDD (Time Division Duplexing), SND (Simultaneous Non-unique Decoding), SWSC (Sliding Window Superposition Coding), MN (Moving Network), D2D (Device-to-Device), mMIMO (massive Multiple Input Multiple Output), VLC (Visible Light Communication), SIC (Self Interference Cancelation), V2X (Vehicle to Everything), UAV (Unmanned Aerial Vehicles).

Table 6. 5G enabling technologies—research directions.

5G Enabling Technologies		Research Gaps and Directions	References
Wide and flexible bandwidth technology	mmWave band communication and large-scale antenna	(1) 3D channel modeling; (2) dynamic power control; (3) user scheduling and congestion control; (4) hardware limitation and adaptive beam-steering technique; (5) design of mobility management and admission control for mmWave-based dense HetNet; (6) design of frequency management schemes for mmWave; (7) Tactile Internet; (8) effective and efficient mmWave implementation in HetNets (access and networking).	[7,87,217]
	Heterogeneous multi-RAT integration	(1) cell-association; (2) traffic-offloading algorithms; (3) interference management schemes in case of inter user and inter cell interference; (4) cross-tier handover, access admission, and mobility management schemes of a multi-tier HetNets.	[196,217]
	Cognitive radio and spectrum sharing	(1) Spectrum sensing (design of cooperative frameworks, choose cooperative secondary users and transmit cooperative information); (2) develop framework and algorithms for group handoff of secondary users and security; (3) simulation of different attacks and scenarios to enhance security; (4) in-depth performance analysis between GFDM and UFMC in CR settings.	[29,110]
Duplexing	In-band full duplexing	(1) Redesign of network and management; (2) antenna and circuit design and development of the theoretical foundation; (3) analyze the throughput of a network of randomly deployment terminals sing stochastic geometry; (4) characterize the capacity advantage due to IFDB in various network scenarios; (5) guidelines to practical design: coding, modulation, power allocation, beamforming, channel estimation, equalization, digital interference cancellation and decoding, (6) design of a MAC layer.	[55,197]
Multiple access and waveform	New waveform	(1) Performance of SIC cancelation or filtering on f-OFDMA; (2) balance of time and frequency dispersion and design an efficient filter prototype for UFMC.	
Access architecture related radio technologies	Advanced small cell	(1) Expect of wireless backhauling on user experience; (2) exploitation of location data and fingerprints in optimizing small cell discovery in terms of time and energy- efficiency; (3) interference management when integrating D2D and small cells.	[34,80]
	Enhanced wireless backhaul	(1) TDD multi-flow coordination schemes to avoid bottlenecks in the downlink backhaul; (2) Backhaul aware association in ultra-dense deployment; (3) reliability and security of the backhaul.	[34,88,198]
	Moving network	(1) Resource allocation and interference in the mobile relay when trains are moving from opposite directions; (2) handover decision of users (more than one train arrive or depart, stop or pass); (3) group mobility for users on board very high-speed vehicles; (4) design od cooperative communication schemes; (5) deployment of moving networks in various vehicle environments, not just on fixes route railways.	[5]
	D2D communication	(1) Interference management (mode selection, resource allocation and power control); (2) integration of novel reputation-based mechanism for identify and avoid malicious users from multiple users in multi-hop D2D communications; (3) testing of D2D interference management schemes in 5G scenarios (mmWave, cell densification).	[29,208,209]
Energy related technologies	Energy harvesting	(1) Improving energy harvesting schemes; (2) simulation of proposed models; (3) integration with other 5G technologies.	[100,182,253]
	UAV	(1) Optimal deployment, mobility and energy-efficient use of UAVs; (2) integration with other 5G technologies.	[81]
Other technologies	mMIMO	(1) Performance of practical mMIMO.	

Legend: IoT (Internet of Things), 5G (Fifth Generation), mmWave (Milimeter-wave), RAT (Radio Access Technology), D2D (Device-to-Device communication), mMIMO (massive Multiple Input Multiple Output), 3D (Three Dimensional), HetNet (Heterogeneous Network), GFDM (Generalized Frequency Division Multiplexing), UFMC (Universal Filtered Multi Carrier), CR (Cognitive Radio), IFDB (In-band Full Duplexing), MAC (Medium Access Control), f-OFDMA (filtered Orthogonal Frequency Division Multiple Access), TDD (Time Division Duplexing), UAV (Unmanned Aerial Vehicle).

5.1. Wide and Flexible Bandwidth Technology

mmWave band communication and large-scale antennas are promising technologies for future 5G networks. The mmWave band covers frequencies from 30 GHz to 300 GHz [202], and from an industry and wireless academia point of view, it is a unique solution for solving 5G capacity requirements [96]. mmWave band communications will provide high data rates [7,202,225–227] utilizing a much larger spectrum bandwidth that can reach up to 5 GHz [39] and by using directional antennas and high attenuation [40]. To provide sufficient antenna gain mmWave requires implementation of large-scale antennas at the transmitter and receiver side [96]. Deployment of large antenna arrays with mmWave will also bring high spectral efficiency, high throughput and channel gain [7]. Despite all of its advantages, mmWave needs line of sight (LOS) operation [41], while the effective communication distance of mmWave signals is within 200 m due to the propagation characteristic of this frequency band [29]. Open problems associated with mmWave include three dimensional (3D) channel modeling, dynamic power control, user scheduling and congestion control, hardware limitation and the adaptive beam-steering technique, as described in [7]. The problem of high power consumption of a large number of antennas in an array [7], high efficiency low complexity adaptive antenna array processing algorithms [29] and innovative hardware architecture of large-scale antenna transmitters [29] still remain unsolved. A proposal of design guidelines in architectures and protocols for mmWave communications is presented in [42]. It is demonstrated that new mmWave technologies, which are under investigation for 5G communications systems, will be able to provide indoor centimeter (cm)-accuracy localization in a robust manner, ideally suited for Assisted Living (AL) [228].

Heterogeneous multi-radio access technologies (multi-RAT) integration is specific to 5G networks radio design that coexists with existing networks. Since 5G networks will not be developed to replace current wireless networks, but rather to advance and integrate existing network infrastructures with new ones [217], we refer to 5G as a heterogeneous network. Multi-RAT is defined as the capability of a mobile network to support multiple radio access technology with seamless interworking among them [229]. When deployed in heterogeneous networks (HetNets), with traffic offloading among different RATs, multi-RAT improves capacity [230], supports better communication rates [34,117–119,231,232], better energy efficiency [34,117–119,231,232], and ensures seamless connectivity with higher QoS [230]. In designing a heterogeneous RAT, researches are finding solutions for cell-association and traffic-offloading algorithms [217].

Cognitive radio with spectrum sharing is a new software defined technology, which is expected to improve the utilization of the congested radio frequency (RF) spectrum [43]. In 5G networks, it is used for designing multi-tier architectures, removing interference among cells, and minimizing energy consumption in the network [44–47,120,233]. Moreover, a spectrum sharing technique can be used along with the CR technology to integrate the 5G spectrum [48,121]. During the practical implementation of the CR and spectrum sensing (SS) technique researchers had to design cooperative frameworks, choose cooperative secondary users and transmit cooperative information during the spectrum sensing, further developing the framework and algorithms for the group handoff of secondary users and enhancing security [29]. Researchers are already working on designs of protocols for different IoT application, based on cognitive radio, and some of them are presented in [49–51,122]. Technologies in this group are used together to optimize 5G performance requirements (e.g., a prototype of mmWave integrated HetNet in [123] and HetNet that incorporates massive multiple-input and multiple-output (mMIMO) and mmWave technologies [217]).

5.2. Advanced Modulation and Coding

This technological group involves advanced modulation (i.e., frequency and quadrature amplitude modulation (FQAM)) and advanced channel coding schemes.

Advanced modulation schemes group includes FQAM, Amplitude and Phase Shift Keying (APSK), Unitary Space-Time Modulation (USTM), Spatial Modulation (SM), Wave Modulation

(WAM), and Orthogonal Time Frequency and Space (OTFS). The FQAM is a combination of frequency shift keying (FSK) and quadrature shift keying (QAM) [124–126]. This modulation can achieve a higher transmission rate for cell edge users reducing interference at cell edge [52,234–236]. FQAM also improves energy efficiency, which makes it adequate for MTC devices with stringent energy consumption requirements [124]. Due to performance advantages in terms of Frame Error Rate (FER) [53,236], FQAM is an ideal candidate for services with high coverage and reliability requirements [237]. APSK is another modulation which draws a lot of attention. It is shown that its main performance gain (i.e., to achieve a channel capacity very close to Shannon's) relies on advanced channel coding and demodulation algorithms [237]. This adds more complexity at the APSK transmitter and receiver. Another modulation, which does not need Channel State Information (CSI) to enable high throughput, is called USTM [237]. USTM and its extension (see [238]) is very useful for 5G services with high mobility [127] or latency and reliability constraints [239]. Other modulation schemes involve SM [128], and proprietary WAM and OTFS [237]. On the basis of the foregoing, researches have analyzed and compared many modulation forms, but still have not found a practical guide on how to choose modulation in any of the 5G use cases.

Advanced channel coding techniques are used for correcting the communication errors caused by noise, interference, and poor signal strength [129]. Authors in [54,129,130] compared turbo, low density parity check (LDPC) and polar codes in decoders in contrast to 5G requirements: (1) maturity; (2) throughput and latency; (3) error correction capability; (4) flexibility; (5) computation complexity; (6) interconnect complexity; (7) high-performance flexible implementation complexity; and (8) backward compatibility. This comparison showed that turbo codes hold the greatest promise for offering high performance throughputs, latencies and error correction capabilities, as well as high degrees of flexibility at the lowest implementation complexity [129] in most 5G use cases. However, further work is needed to implement a decoder based on this code followed by detail analysis.

5.3. Duplexing

This technological group includes in-band full duplexing (FD) and dynamic time division duplexing (TDD).

In-band FD or simultaneous data transmission and reception will provide a 1000-fold increase in throughput [131,240], double spectral efficiency [132,240], and reduce the air interface delay [240]. The central research problem for the practical implementation of an in-band full-duplex radio is the attenuation of the self-interference signal by an adequate amount [132]. For the practical implementation of an in-band full-duplex radio, many aspects of network design and management need to be restructured, where terminals antenna and circuit design and the development of theoretical foundation are in focus [55]. One practical implementation of the full-duplex radio is shown in [133]. Authors in [134] commented that for the design of a full-duplex radio, it is necessary to unify researches from three domains, i.e., RF circuit and system design, digital signal processing and networking.

Dynamic TDD is the predecessor of FD transmission technology and a candidate for 5G [133,135]. It represents a scheduling technique in which every base station (BS) is free to choose its own uplink/downlink (UL/DL) split [136–139]. This technique is used to adapt the allocation of network resources to variable traffic requirements [139,140], often found in ultra-densely deployed networks [141]. The dynamic TDD can significantly increase bandwidth efficiency [142] and provide higher throughput and low latency [141]. However, it is characterized by the severe co-channel interference (CCI) [141–143]. Dynamic TDD could be used in combination with D2D communication and in the self-backhauling scenario, as described in [55].

5.4. Multiple Access and Waveforms

Multiple access techniques are becoming an important technology in 5G because of their ability to support mMTC activity-based IoT classes with urgent deploy demand [144]. They include several non-orthogonal multiple access (NOMA) forms: multi-user shared multiple access (MUSA) [145–148],

resource spread multiple access (RSMA) [149], sparse code multiple access (SCMA) [150–152], pattern division multiple access (PDMA) [153–155], interleave-division multiple access (IDMA) [156,157], and NOMA by power domain [158]. The NOMA is a radio access technology design for enabling greater spectrum efficiency [56,159–162,207], higher cell-edge throughput, relaxed channel feedback, and low transmission latency [99]. NOMA can be employed to enhance user fairness and to support massive connections with diverse QoS requirements [163]. In NOMA, there are several challenges and open issues, such as dynamic user pairing, the impact of transmission distortion, the impact of interference, resource allocation, NOMA with multiple antennas, heterogeneous networks, outage probability analysis, practical channel model, uniform fairness, NOMA with antenna selection, carrier aggregation and other challenges as discussed in [216]. There are still many challenging issues for SCMA, which need to be solved in future work. For example, there are several open issues in SCMA transceiver design [57], the optimization of algorithms for user grouping and power allocation [164], and further enhancement of the SCMA and MIMO combination.

New waveforms have become a serious candidate for 5G being studied in terms of: (1) modulation based on pulse shaping: filter bank multicarrier (FBMC) [58], generalized frequency division multiplexing (GFDM) [59], pulse shaped OFDM [165] and QAM-FBMC [241]; (2) modulation based on sub-band filtering: universal filtered multi carrier (UFMC) [242], filtered OFDM (f-OFDM) [243] and resource block f-OFDM (RB-f-OFDM) [244]; other modulation format: guard interval discrete Fourier transform spread OFDM (GI DFT-s-OFDM) [245], spectrally-precoded OFDM (SP-OFDM) [246] and orthogonal time frequency and space (OTFS) [247]. The f-OFDM is seen as a potential candidate for IoT applications [60]. Due to narrow sub-bands and thereby pure detection performance, additional processing is needed [60]. Performance of self-interference cancellation (SIC) or filtering on f-OFDMA [60] remains a topic for further study. Additionally, it is interesting to consider the balance of time and frequency dispersion in UFMC, as well as the design of an efficient prototype filter according to application scenarios [60].

5.5. Advanced Interface Management

This technological group includes receiver advances in terms of simultaneous non-unique decoding (SND) and sliding-window superposition coding (SWSC).

SND follows a rule that implies that each receiver attempts to recover the code words from intended and interfering senders [248]. The combination of advanced receivers and joint scheduling provides an improvement of over 50% in cell edge throughput without sacrificing the cell average throughput [5]. This gain demonstrates that if 5G networks incorporate advanced interference management, they will provide a virtually edgeless end-user experience [5].

SWSC [61,62] combines the theory concept from superposition coding without rate splitting [63], block Markov coding [166,167], successive cancellation decoding [63,64,168], and sliding-window decoding [65,169,249]. The sliding-window coded modulation (SWCM) aims to mitigate inter cell interference at the physical layer by achieving simultaneous decoding performance with point-to-point channel codes, low-complexity decoding, and minimal coordination overhead [170]. The realization of the theoretical concept to the practical transmission coding scheme is an important research direction in SWSC [5].

5.6. Access Architecture Related Radio Technology

This technological group includes ultra-densification, enhanced wireless backhaul, moving networks, and D2D communication.

Advanced small cell [66,123] deployment is considered to be one of the key enablers for achieving many of the requirements currently envisioned for 5G [67], with higher data rates [68,171,250] and more capacity [171–174], as most important. Higher data rates and smaller battery consumption can be achieved using short distances in the small cell [69]. Solving the capacity and data rate challenge with network densification could be very expensive in terms of equipment, maintenance,

and operations [200]. With denser cell deployment come significant challenges in the design of a high-performance backhauling system for RAN and the impact of backhaul on radio resource management (RRM) [70]. However, there are in existence many interesting topics to be further investigated which can be grouped around several problem areas [34]: user association, backhauling, interference management, energy efficiency, and propagation modelling.

Moving networks (MN) are the combination of multi-hop and vehicular communications concepts [69,71]. The MN deploys one or several moving relay node(s) (MRNs) [175,176,251] on vehicles that form their own cell(s) inside the vehicle to serve vehicular users [72]. Challenges in using MRN are efficient backhauling, design of efficient resource allocation and interference management techniques, as well as mobility management schemes to exploit the benefit of group handovers for vehicular user equipment (UE) devices served by the same MRN [72]. The main research challenges related to MN concept are associated with complexity management due to the mobility of access points and providing a high-rate wireless backhaul link from the moving cell to the fixed network [69].

Enhanced wireless backhaul is one of the main challenges in hyper-dense 5G networks [34,217]. The performance of wireless backhaul is dictated by the environment and traffic profile of the intended use case [9]. In UDN except mmWave, new approaches are needed, e.g., interface aware routing and intelligent resource allocation [9], as these links may occupy part of the spectrum used in the access network [73]. Latency and reliability of this link are important issues to be considered being prone to blocking and fading [69]. In any case, further research is necessary to explore TDD multi-flow coordination schemes to avoid bottlenecks in the downlink backhaul [9]. The guidelines for deploying future 5G wireless backhaul networks in economical and highly energy-efficient ways are provided in [74].

D2D communication is defined as a direct route of data traffic between spatially closely located mobile UE [75,177,178]. The advantages of D2D communication compared with the traditional cellular method include the reduction of transmission latency [73,75] and power consumption [75], improvements in coverage [71,76,179], spectral and energy efficiency [77], and throughput, when power control and resource allocation methods are used [78]. The use of D2D communications has an overall positive impact on system capacity in cellular environments. D2D communication brings challenges and complexities related to interference management [79,80], QoS requirements [79], and resource allocation [34,215]. The design of D2D direct communication link is still a hot topic [29].

5.7. Energy Related Technologies

IoT applications span a broad range of domains including home automation, healthcare, surveillance, transportation, smart environments, etc. One of the most important obstacles for implementing such an impressive scheme is supplying adequate energy to operate the network in a self-sufficient way without compromising QoS [180]. Since energy efficiency is of most importance to battery constrained IoT devices, researches have focused their work on the development of the device energy saving mechanisms.

Energy harvesting is a new paradigm which uses solar, thermal, wind and kinetic energy sources to power sensor nodes and consequently prolong network lifetime [181]. Among different energy harvesting methods, wireless energy harvesting (WEH) has proven to be one of the most promising solution for energy aware IoT devices, because of its simplicity, ease of implementation, and availability [180]. This concept will improve some of the 5G high network communication requirements, like reliability of IoT communications, as presented in [182]. Authors in [252] have shown that the integration of social awareness and energy harvesting in D2D communication results in higher system capacity. Network architectures proposed in [100,253] have shown better energy efficiency in 5G wireless network. Researchers could work on improving this architecture and integration with various 5G technologies in the future.

UAV implies a flying vehicle without the driving presence of a human pilot to control it on-board [113]. Drones are a possible example of UAVs, although many other robotics related

applications could be part of this application group [113]. UAVs can play a vital role in IoT scenarios where devices are unable to transmit over a long distance due to their energy constraints. In this case UAVs play the role of moving aggregators which fly toward one IoT devices, collect the data, and transmit it to other devices [81]. In [183] authors presented UAV–based floating relay (FR) for cell dynamic and coverage improvement. Deployment of unmanned aerial base stations in 5G heterogeneous architecture can improve throughput [254], coverage [183–185,254], connectivity [184,185], and 5th percentile spectral efficiency [254]. On-demand wireless systems with low-altitude UAVs are in faster to deploy, more flexibly reconfigured, and likely to have better communication channels due to the presence of short-range line-of-sight links [186] when compared to terrestrial communications or those based on high-altitude platforms. The utilization of highly mobile and energy-constrained UAVs for wireless communication introduces many new challenges [186], some of them are listed in [82]. To effectively use UAVs for IoT, several challenges must be addressed such as optimal deployment, mobility and energy-efficient use of UAVs [81].

5.8. Other Technologies

Along with the aforementioned technologies, a lot of research attention is devoted to technologies, such as mMIMO, visible light communication (VLC), SIC, vehicle-to-everything (V2X), etc.

mMIMO technology [73,123] promises significant gains in data rate and link reliability [187], reduces latency and energy [13], simplifies media access control (MAC) layer [13], shows robustness against intentional jamming [13], unintended man-made interface [83], and increases capacity [127,198] due to spatial multiplexing [123]. Reliable links are provided by benefiting from spatial diversity and the mitigating effects of fast fading, beamforming, and zero forcing caused by multi-user interference [203]. The mMIMO can be exploited to extend the coverage of higher frequency bands by relying on beamforming gains [200]. Other specific benefits of mMIMO system are: increased capacity 10 times or more with simultaneous improvement in radiated energy efficiency in the order of 100 times, the possibility to use inexpensive, low-power components, the reduction of latency on the air interface, and multiple access layer simplification [83]. The low complexity mMIMO uplink scheme for IoT lightweight devices is presented in [188]. The performance of practical mMIMO system needs to be investigated, since the research community pays attention to analyzing the performance of mMIMO in an ideal channel state information assumption [73].

VLC is a growing technology for short range, high capacity LOS optical links [189]. It uses a visible range of the electromagnetic spectrum (370–780 nm) which provides data transmission and room illumination using light emitting diodes (LEDs). Prominent features of VLC are an abundant license-free spectrum, the ability to provide multiple gigabit-per-second data rates, low energy consumption, and low implementation costs [190]. VLC is sensitive to sunlight and is not able to work long range without LOS. Since VLC coverage is LOS limited [84,191], this is known as LOS blocking [192]. With poor performance in non-line-of-sight scenarios, VLC networks fail to provide convenient UL coverage at the current state-of-the-art, and each AP illuminates only a small confined cell compared to cellular RF networks [192]. VLC applications are expected to include IoT, wireless Internet access, and vehicle-to-vehicle (V2V) communications, broadcast from LED, M2M communications, positioning systems, and navigation. The problem of how to provide mobile applications over VLC is still quite an open one [189].

SIC implementation enables low-latency applications in a cost effective manner [132]. SIC can be used to increase link capacity, spectrum virtualization, any-division duplexing, novel relay solutions, and enhanced interface coordination [132]. Practical SIC implementation is still in the investigation phase.

V2X is the communication between a vehicle and everything with which it might interact (e.g., other vehicles, traffic operators and service providers). It is used in conjunction with D2D for coverage extension and latency reduction in 5G networks [255]. V2X communication might play a vital role in terms of implementing smart and efficient traffic solutions [193], improving road safety, traffic efficiency,

and driver convenience. To reduce traffic congestion one may use an intelligent route management system based on V2X communication [7]. Since it has big potential in autonomous driving application, the localization of vehicles in dense urban areas is difficult, so several algorithms are proposed to solve the problems [194]. Many V2X applications are described in existing literature [85,195]. However, there is still a need to design vehicular mobility management strategies in next UDN networks and to find a way to enhance existing connected vehicle services with V2X [86].

These technologies are highlighted as the main 5G driver in existing literature [5–7,96,207]. Authors compare these technologies and try to implement them through prototypes, or test the combination of these technologies. However, other technologies exist, such as cloud-RAN [7], Coordinated Multi-Point (CoMP), White Space Spectrum [208], which are not in the scope of this paper.

5.9. Summary of IoT in Enabling Technologies

Performance requirements of activity-based IoT application classes have served to identify the enabling technologies in the radio access part of 5G networks, which are summarized in Table 5. Various 5G enabling technologies have been sorted into eight technological groups:

1. Wide and flexible bandwidth technology: mmWave band communication and large-scale antennas, heterogeneous multi-RAT integration, cognitive radio, and spectrum sharing;
2. Advanced modulation and coding: advanced modulation schemes, advanced channel coding;
3. Duplexing: in-band FD, dynamic TDD;
4. Multiple access and waveform: multiple access, new waveforms;
5. Advanced interface management: SND, SWSC;
6. Access architecture related radio technologies: advanced small cell, MN, enhanced wireless backhaul, D2D;
7. Energy related technologies: energy harvesting, UAV;
8. Other technologies: mMIMO, VLC, SIC, V2X.

These technologies have been discussed in terms of their possibilities to meet performance requirements of activity-based IoT application classes and identify research gaps and directions, which are summarized in Table 6.

We have highlighted technologies only used to satisfy high priority performance requirements of mMTC activity-based IoT application classes, i.e., connection density, positioning accuracy, and coverage. We have focused on these technologies from telco's perspective, since the area of mMTC activity-based IoT classes is more mature, as its development has already started within LTE, while uMTC poses new research questions to be answered in order to achieve unprecedented levels of reliability needed for new applications in 5G.

The new waveform technology may be used to fulfil the requirements of almost every activity-based IoT application class in terms of connection density. This technology affects the connection density by definition. Analysis of different modulation formats indicates that f-OFDM is the most suitable candidate for IoT applications with high connection density requirements. In this regard, a deeper and more comprehensive analysis of f-OFDM is needed in the context of different mMTC activities (i.e., ticketing, tracking, monitoring, and managing/controlling). The result of the analysis to be performed should be the discovery of an optimal waveform for each activity-based IoT class which has to be integrated with other 5G service classes.

Large scale antennas and advanced small cell technologies can be used to satisfy the requirements of ticketing-, tracking-, and uMTC managing/controlling-based IoT application classes in terms of positioning accuracy. Fulfilling these performance requirements is rather a business than a technical decision. The size of the 5G cell must be planned to meet the desired position accuracy since it is very important for future IoT applications, such as automated driving, remote surgery, robotics, taxis, etc.

Wide and flexible bandwidth technologies, in-band full duplex technologies, access architecture related radio technologies, and mMIMO can be used to achieve the high requirements of ticketing-,

tracking-, and monitoring-based IoT application classes in terms of coverage. The combination of these technologies should be tested to find the optimal solution to meet these performance requirements. Since small cell deployment has an impact on required position accuracy, they should serve as a basis for achieving the desired level of coverage. Along with small cell deployment, mMIMO and mmWave technologies are present in almost every 5G prototype.

Since this section discusses radio technologies used to accomplish performance requirements of mMTC activity-based IoT application classes, it can serve telecom operators and other interested parties depending on their interest and potential to utilize these findings towards IoT implementation and monetization. It describes each technology in terms of definition, advantages, disadvantages, and the possible impact on the performance requirements of a broad range of IoT applications that drive the deployment of 5G networks. Moreover, this section can be useful to the research community interested in this attractive field to address recognized research gaps and directions.

6. Discussion

Telecom operators have the opportunity to capture a share of the revenue that is generated by IoT implementation depending on their role in the IoT value chain. Regardless of their role, telcos have to face many technical challenges in order to meet the changing connectivity and performance requirements. Since current cellular networks limit numerous IoT applications, new technologies are being introduced leading towards 5G networks. Therefore, telcos have to focus on deploying these network technologies in order meet the changing requirements necessary to achieve success in IoT.

In this regard, writing this paper was motivated by the challenge of providing an enhanced understanding of the scope and classification of the broad range of IoT applications in order to determine and prioritize their performance requirements with the goal of specifying the enabling technologies towards 5G networks. The aim has been to propose a new classification of IoT applications, define and prioritize the performance requirements of such IoT application classes, and give insight into state-of-the-art technologies used to meet these requirements from telco's point of view. The motivation that led to the focus of this paper being on IoT applications classification, performance requirements, and 5G enabling technologies could be explained by telco's need for added value from IoT services. Hence, an analysis of IoT customer service requests inspired us to propose an activity-based classification of IoT applications as it can, according to authors' best knowledge, has allowed telcos to more precisely specify their performance requirements and 5G enabling technologies, thereby improving customer experience.

In this regard, the paper fulfilled the following four objectives: (1) the identification of activities relevant to IoT applications and their usage as a new criterion for IoT application classification; (2) the specification and prioritization of performance requirements of such IoT applications classes; (3) the analysis of the radio technologies used to accomplish IoT application requirements; and (4) the identification of the research gaps and the recommendation of new research directions.

Through fulfilling these objectives, we reviewed literature from the fields of IoT in 5G service classifications, IoT in 5G performance requirements, and IoT in 5G enabling technologies. Since the first aim of this paper was to propose a new classification of IoT applications, the existing approaches have been summarized to serve as a basis to identify their drawbacks and formulate the appropriate solution to overcome them. In this context, we proposed a new approach to IoT applications classification, which was based on the activity as a new classification criterion denoting the main function/purpose of specific IoT application. This approach enabled a clearer and more precise positioning of particular application in the IoT application spectrum, as well as the determination of performance requirements and enabling technologies from telco's point of view. The activity-based IoT application classification facilitated the specification of the performance requirements and determination of technologies which enable these requirements to be fulfilled. In this context, the proposed approach served as a basis for the simplification of the realization of particular IoT application. For example, although the monitoring- and tracking-based applications seem to be similar, they set different performance requirements on the

network, thereby allowing the proposed approach to precisely grade the IoT applications from the same IoT application domain.

Furthermore, we associated the activity-based IoT application classes with the 5G service classes, i.e., mMTC and uMTC, in order to specify and prioritize their performance requirements. It was determined that almost each activity-based IoT application class poses high requirements in terms of connection density, whereas ticketing-based and tracking-based IoT application classes additionally require high positioning accuracy and coverage. In order to analyze these performance requirements, we have summarized the radio technologies used to implement the future 5G networks. The focus was on the technologies used to meet the aforementioned performance requirements of mMTC activity-based IoT application classes, since mMTC is more mature, as its deployment has already started, while uMTC requires further research to achieve incomparable levels of reliability needed to enable new applications in 5G.

The analysis showed that new waveform technology can be used to meet the requirements in terms of connection density; large-scale antennas and advanced small cell technologies can be used for the purpose of satisfying the requirements in terms of positioning accuracy; while wide and flexible bandwidth technologies, in-band full duplexing technologies, and access architecture related radio technologies (i.e., advanced small cell, enhanced wireless backhaul, moving network, and D2D communication) can be used to achieve high requirements in terms of coverage. In addition, the conducted research study allowed us to highlight a number of open research issues that could serve the research community and other stakeholders interested in this contemporary and attractive field. In this context, we have recognized research gaps and directions which mostly relate to network redesign and optimization in order to accommodate large-scale IoT applications.

7. Conclusions

The IoT paradigm has the potential to revolutionize all areas of daily life of individuals, businesses, and society as a whole. Telcos enjoy a central role in the paradigm of IoT because of owing communication infrastructure which is exposed to the numerous technical challenges due to changing connectivity and performance requirements of various IoT applications. These requirements cannot be met with the current cellular networks which create the need to introduce new types of technologies leading toward 5G networks as the main driver for enabling numerous IoT applications. In this context, this paper proposed the activity-based classification of IoT applications and specification of their performance requirements in order to identify 5G radio technologies used to meet them, all from telco's perspective. Activity-based classification of IoT applications indirectly allowed the construction and distribution of performance, as IoT value proposition, while using 5G enabling technologies, as IoT resources in telco's internal value chain. In this context, high performance requirements of each activity-based IoT application class (i.e., connection density, positioning accuracy, coverage) served as a basis to analyze various enabling technologies in radio access part of 5G network in terms of advantages, disadvantages, and research gaps.

On the basis of the conducted analysis, we concluded that the following technologies can meet the high performance requirements of mMTC activity-based IoT application classes: new waveform technology (in terms of connection density), large scale antennas (in terms of positioning accuracy), wide and flexible bandwidth technologies, in-band full duplexing technologies, and access architecture related radio technologies (in terms of coverage). According to the identified research gaps, one may conclude that the optimal solutions for each technology are still in its infancy. This implies that any 5G enabling technology should be first implemented and tested through prototype construction before the deployment of specific IoT application. Thereafter, the technology combination that meets the performance requirements of each activity-based IoT class should first be found. Finally, the practical implementation of different technological combinations may further lead to a deep and comprehensive analysis of QoS and QoE in the context of IoT applications.

References

1. Antonopoulos, A.; Perillo, C.; Verikoukis, C. Internet Service Providers vs. Over-the-Top Companies: Friends or Foes?—Short talk. *SIGMETRICS Perform. Eval. Rev.* **2016**, *44*, 37. [CrossRef]

2. Antonopoulos, A.; Kartsakli, E.; Perillo, C.; Verikoukis, C. Shedding Light on the Internet: Stakeholders and Network Neutrality. *IEEE Commun. Mag.* **2017**, *55*, 216–223. [CrossRef]

3. IoT-Ignite. Role of Telcos in Internet of Things. Available online: https://devzone.iot-ignite.com/wp-content/uploads/2017/01/Role-of-Telcos-in-IoT.pdf (accessed on 5 September 2017).

4. Kaur, S.; Singh, I. A Survey Report on Internet of Things Applications. *Int. J. Comput. Sci. Trends Technol.* **2016**, *4*, 330–335.

5. 5G Forum. 5G Vision, Requirements, and Enabling Technologies. Available online: http://kani.or.kr/5g/whitepaper/5G%20Vision,%20Requirements,%20and%20Enabling%20Technologies.pdf (accessed on 10 July 2017).

6. Saha, R.K.; Saengudomlert, P.; Aswakul, C. Evolution towards 5G Mobile Networks—A survey on Enabling Technologies. *Eng. J.* **2016**, *20*, 87–112. [CrossRef]

7. Akyildiz, I.F.; Nie, S.; Lin, S.C.; Chandrasekaran, M. 5G roadmap: 10 Key Enabling Technologies. *Comput. Netw.* **2014**, *106*, 17–48. [CrossRef]

8. NGMN Alliance. 5G White Paper-Executive Version. Available online: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf (accessed on 10 July 2017).

9. Popovski, P.; Mange, G.; Gozalves-Serrano, D.; Rosowski, T.; Zimmermann, G.; Agyapong, P.; Fallgren, M.; Hoglund, A.; Queseth, O.; Tullberg, H.; et al. Deliverable D6.6 Final Report on the METIS 5G System Concept and Technology Roadmap. Available online: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D6.6_Summary.pdf (accessed on 10 July 2017).

10. Aly, H.; Elmogy, M.; Barakat, S. Big Data on Internet of Things: Applications, Architecture, Technologies, Techniques, and Future Directions. *Int. J. Comput. Sci. Eng.* **2015**, *4*, 300–313.

11. Vermesan, O.; Friess, P. *Internet of Things—From Research and Innovation to Market Deployment*; Riverbed: Aalborg, Denmark, 2014.

12. Amaral, L.A.; de Matos, E.; Tiburski, R.T.; Hessel, F.; Lunardi, W.T.; Marczak, S. Middleware Technology for IoT Systems: Challenges and Perspectives Toward 5G. In *Internet of Things (IoT) in 5G Mobile Technologies*; Mavromoustakis, C., Mastorakis, G., Batalla, J., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 333–367.

13. OECD. Machine-to-Machine Communications: Connecting Billions of Devices. Available online: http://www.oecd-ilibrary.org/docserver/download/5k9gsh2gp043-en.pdf?expires=1499694889&id=id&accname=guest&checksum=F403637E723693581F03EA9D6F3A3C3B (accessed on 10 July 2017).

14. Zheng, K.; Hu, F.; Wang, W.; Xiang, W. Radio Resource Allocation in LTE Advanced Cellular Networks with M2M Communications. *IEEE Commun. Mag.* **2012**, *50*. [CrossRef]

15. Zhang, Q.; Fitzek, F.H.P. Mission Critical IoT Communication in 5G. In *Future Access Enablers for Ubiquitous and Intelligent Infrastructures*; Springer International Publishing: Cham, Switzerland, 2015; Volume 159, pp. 35–41.

16. Oughton, E.J.; Frias, Z. Exploring the Cost, Coverage and Rollout Implications of 5G in Britain. Available online: http://www.itrc.org.uk/wp-content/uploads/Exploring-costs-of-5G.pdf (accessed on 10 July 2017).

17. Chen, Y.; Wang, W. Machine-to-Machine Communication in LTE-A. In Proceedings of the 2010 IEEE Vehicular Technology Conference (VTC 2010-Fall), Ottawa, ON, Canada, 6–9 September 2010.

18. Booysen, M.J.; Gilmore, S.J.; Zeadally, S.; van Rooyen, G.J. Machine-to-Machine (M2M) Communications in Vehicular Networks. *KSII Trans. Internet Inf. Syst.* **2012**, *6*, 529–546. [CrossRef]

19. Niyato, D.; Xiao, L.; Wang, P. Machine-To-Machine Communications for Home Energy Management System in Smart Grid. *IEEE Commun. Mag.* **2011**, *49*, 53–59. [CrossRef]

20. De Matos, W.D.; Gondim, P.R.L. M-Health Solutions Using 5G Networks and M2M Communications. *IT Prof.* **2016**, *18*, 24–29. [CrossRef]

21. Investopedia. Business Model. Available online: http://www.investopedia.com/terms/b/businessmodel.asp (accessed on 3 September 2017).

22. Gassmann, O.; Frankenberger, K.; Csik, M. Revolutionizing the Business Model. In *Management of the Fuzzy Front End of Innovation*; Springer: New York, NY, USA, 2014.

23. Vermesan, O.; Bahr, R.; Gluhak, A.; Boesenberg, F.; Hoeer, A.; Osella, M. D02.01 IoT Business Models Framework. Available online: http://www.internet-of-things-research.eu/pdf/D02_01_WP02_H2020_ UNIFY-IoT_Final.pdf (accessed on 3 September 2017).

24. Dijkmana, R.M.; Sprenkelsa, B.; Peetersa, T.; Janssenb, A. Business Models for the Internet of Things. *Int. J. Inf. Manag.* **2015**, *35*, 672–678. [CrossRef]

25. Farooq, M.O.; Sreenan, C.J.; Brown, K.N.C. Research Challenges in 5G Networks: A HetNets Perspective. In Proceedings of the 19th International ICIN Conference Innovations in Clouds, Internet and Networks, Paris, France, 1–3 March 2016.

26. Hu, F. 5G Overview: Key Technologies. In *Opportunities in 5G Networks*, 1st ed.; Hu, F., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 1–557.

27. Ford, R.; Zhang, M.; Mezzavilla, M.; Duttam, S.; Rangap, S.; Zorzi, M. Achieving Ultra-Low Latency in 5G Millimeter Wave Cellular Networks. *IEEE Commun. Mag.* **2017**, *55*, 196–203. [CrossRef]

28. Prasad, R. IMT for 2020 and Beyond. In *5G Outlook—Innovations and Applications*; Riverbed Publishers: Gistrup, Denmark, 2016; pp. 1–260.

29. Ma, Z.; Zhang, Z.; Ding, Z.; Fan, P.; Li, H. Key Techniques for 5G Wireless Communications: Network Architecture, Physical Layer, and MAC Layer Perspectives. *Sci. China Inf. Sci.* **2015**, *50*, 1–20. [CrossRef]

30. Rappaport, T.S.; Daniels, R.C.; Heath, R.W.; Murdock, J.N. Introduction. In *Millimeter Wave Wireless Communications*; Pearson Education: Upper Saddle River, NJ, USA, 2014.

31. Ge, X.; Chen, J.; Ying, S.; Chen, M. Energy and Coverage Efficiency Trade-off in 5G Small Cells Network. *arXiv* **2016**, arXiv:1612.04459.

32. Hossain, S. 5G Wireless Communication Systems. *Am. J. Eng. Res.* **2013**, *2*, 344–353.

33. Artiga, X.; Nunez-Martinez, J.; Perez-Neira, A.; Lendrino Vela, G.J.; Fare Garcia, J.M.; Ziaragkas, G. Terrestrial-satellite Integration in Dynamic 5G Backhaul Networks. In Proceedings of the 2016 8th Advanced Satellite Multimedia Systems Conference and the 14th Signa (ASMS/SPSC), Palma de Mallorca, Spain, 5–7 September 2016.

34. Kamel, M.; Hamouda, W.; Youssef, A. Ultra-Dense Networks: A Survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2522–2545. [CrossRef]

35. Kartsakli, E.; Lalos, A.S.; Antonopoulos, A.; Tennina, S.; Di Renzo, M.; Alonso, L.; Verikoukis, C. A Survey on M2M Systems for mHealth: A Wireless Communications Perspective. *Sensors* **2014**, *10*, 18009–18052. [CrossRef] [PubMed]

36. Morris, A. DT, Chinese Operators take NB-IoT to Market. Available online: http://www.lightreading.com/ iot/nb-iot/dt-chinese-operators-take-nb-iot-to-market/d/d-id/734022 (accessed on 6 September 2017).

37. Vermesan, O.; Eisenhauer, M.; Sunmaeker, H.; Guillemin, P.; Serrano, M.; Tragos, E.Z.; Valino, J.; van der Wees, A.; Gluhak, A.; Bahr, R. Internet of Things Cogitive Transformation Technology Research Trends and Applications. In *Cognitive Hyperconnected Digital Transformation*; Vermesan, O., Bacquet, J., Eds.; River Publishers: Delft, The Netherlands, 2017; pp. 17–95.

38. Gawas, A.U. An Overview on Evolution of Mobile Wireless Communication Networks: 1G–6G. *Int. J. Recent Innov. Trends Comput. Commun.* **2015**, *3*, 3130–3133. [CrossRef]

39. Qi, Y.; Hunukumbure, M.; Nekovee, M.; Sgardoni, V. Quantifying Data Rare and Bandwith Requirements for Immersive 5G Experience. In Proceedings of the 2016 IEEE International Conference on Communications Workshops, Kuala Lumpur, Malaysia, 23–27 May 2016.

40. Sim, G.H.; Asadi, A.; Loch, A.; Widmer, J. Opp-Relay: Managing Directionaly and Mobility Issues of Millimeter-Wave via D2D Communication. In Proceedings of the I9th International Conference on COMSNETS 2017, Bangalore, India, 4–8 January 2017.

41. Mer, S.B. Smart Vehicle-to-Vehicle Communication with 5G Technology. *Int. J. Recent Innov. Trends Comput. Commun.* **2015**, *3*, 3241–3244.

42. Niu, Y.; Li, Y.; Jin, D.; Su, L.; Vasilakos, A.V. A Survey of Millimeter Wave Communications (Mmwave) for 5G: Opportunities and Challenges. *Wirel. Netw.* **2016**, *21*, 2657–2676. [CrossRef]

43. Rusek, F.; Persson, D.; Lau, B.K.; Larsson, E.G.; Marzetta, T.L.; Edfors, O.; Tufvesson, F. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Process. Mag.* **2012**, *30*, 1053–5888. [CrossRef]

44. Hong, X.; Wang, J.; Wang, C.X.; Shi, J. Cognitive Radio in 5G: A Perspective on Energy-Spectral Efficiency Trade-Off. *IEEE Commun. Mag.* **2014**, *52*, 46–53. [CrossRef]

45. Huang, L.; Zhu, G.; Du, X. Cognitive Femtocell Networks: An Opportunistic Spectrum Access for Future Indoor Wireless Coverage. *IEEE Wirel. Commun.* **2013**, *20*, 44–51. [CrossRef]

46. Lee, C.H.; Shih, C.Y. Coverage Analysis of Cognitive Femtocell Networks. *IEEE Wirel. Commun. Lett.* **2014**, *3*, 177–180. [CrossRef]

47. Wang, W.; Yu, H.; Huang, A. Cognitive Radio Enhanced Interference Coordination for Femtocell Networks. *IEEE Commun. Mag.* **2013**, *51*, 37–43. [CrossRef]

48. Menon, R.; Buehrer, R.M.; Reed, J.H. On the Impact of Dynamic Spectrum Sharing Techniques on Legacy Radio Systems. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 4198–4207. [CrossRef]

49. Zhang, Y.; Yu, R.; Nekovee, M.; Liu, Y.; Xie, S.; Gjessing, S. Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid. *IEEE Netw.* **2012**, *26*, 6–13. [CrossRef]

50. Aijaz, A.; Aghvami, A.H. Cognitive Machine-To-Machine Communications for Internet-of-Things: A Protocol Stack Perspective. *IEEE Internet Things J.* **2015**, *2*, 103–112. [CrossRef]

51. Li, R.; Lu, B.; McDonald-Maier, K.D. Cognitive Assisted Living Ambient System: A Survey. *Digit. Commun. Netw.* **2015**, *1*, 229–252. [CrossRef]

52. Andrews, J.G.; Buzzi, S.; Choi, W.; Hanly, S.V.; Lazano, A.; Soong, A.C.K.; Zhang, J.C. What Will 5G Be? *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1065–1082. [CrossRef]

53. Samsung Electronics. 5G Vision, Version 2. Available online: http://www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-0.pdf (accessed on 10 July 2017).

54. Iscan, O.; Lentner, D.; Xu, W. A Comparison of Channel Coding Schemes for 5G Short Message Transmission. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–8 December 2016.

55. Sabharwal, A.; Schniter, P.; Guo, D.; Bliss, D.W.; Rangarajan, S.; Wichman, R. In-Band Full-Duplex Wireless: Challenges and Opportunities. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1637–1652. [CrossRef]

56. Ding, Z.; Liu, Y.; Choi, J.; Sun, Q.; Elkashlan, M.; Chin-Lin, I.; Vincent Poor, H. Application of Non-Orthogonal Multiple Access in LTE and 5G Networks. *IEEE Commun. Mag.* **2006**, *55*, 185–191. [CrossRef]

57. Lei, W. Non-Orthogonal Multiple Access Schemes for Future Cellular Systems. Ph.D. Thesis, University of Surrey, Surrey, UK, 17 May 2017.

58. Farhang-Boroujeny, B. OFDM Versus Filter Bank Multicarrier. *Signal Process. Mag.* **2011**, *28*, 92–112. [CrossRef]

59. Bellanger, M. FS-FBMC: An Alternative Scheme for Filter Bank Multicarrier Transmission. In Proceedings of the 2012 5th International Symposium on Communications Control and Signal Processing, Rome, Italy, 2–4 May 2012.

60. Cai, X.; Qin, Z.; Cui, F.; Li, G.Y.; McCann, J.A. Modulation and Multiple Access for 5G Network. Available online: https://arxiv.org/pdf/1702.07673.pdf (accessed on 10 July 2017).

61. Wang, L.; Sasoglu, E.; Kim, Y.H. Filtered OFDM: Sliding-Window Superposition Coding for Interference Networks. In Proceedings of the International Symposium Information Theory, Honolulu, HI, USA, 29 June–4 July 2014.

62. Wang, L. Channel Coding Technique for Network Communication. Ph.D. Thesis, University of California, San Diego, CA, USA, December 2015.

63. Cover, T.M. Broadcast Channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 2–14. [CrossRef]

64. Liao, H.H.J. Multiple Access Channels. Ph.D. Thesis, University of Hawaii, Honolulu, HI, USA, September 1972.

65. Kramer, G.; Gastpar, M.; Gupta, P. Cooperative Strategies and Capacity Theorems for Relay Networks. *IEEE Trans. Inf. Theory* **2005**, *51*, 3037–3063. [CrossRef]

66. Bhushan, N.; Li, J.; Malladi, D.; Gilmore, R.; Brenner, D.; Damnjanovic, A.; Sukhavasi, R.; Patel, C.; Geirhofer, S. Network Densification: The Dominant Theme for Wireless Evolution into 5G. *IEEE Commun. Mag.* **2014**, *52*, 82–89. [CrossRef]

67. Chin, W.H.; Fan, Z.; Haines, R. Emerging Technologies and Research Challenges for 5G Wireless Networks. *IEEE Wirel. Commun.* **2014**, *21*, 106–112. [CrossRef]

68. Biral, A.; Centenaro, M.; Zanellan, A. The Challenges of M2M Massive Access in Wireless Cellular Networks. *Digit. Commun. Netw.* **2015**, *1*, 1–19. [CrossRef]

69. Barreto, A.N.; Faria, B.; Almeida, E.; Rodriguez, I.; Lauridsen, M.; Amorim, R.; Vieira, R. 5G-Wireless Communications for 2020. *J. Commun. Inf. Syst.* **2016**, *31*, 146–163. [CrossRef]

70. Wang, N.; Hossain, E.; Bhargava, V.K. Backhauling 5G Small Cells: A Radio Resource Management Perspective. *IEEE Wirel. Commun.* **2015**, *22*, 41–49. [CrossRef]

71. Chen, L.; Huang, Y.; Xie, F.; Gao, Y.; Chu, H.; Li, Y.; Liang, F.; Yuan, Y. Mobile Relay in LTE-Advanced Systems. *IEEE Commun. Mag.* **2013**, *51*, 144–151. [CrossRef]

72. Svensson, T. Moving Cells/Networks. Available online: http://snow.itn.liu.se/asbtr/SNOWTommySvensson.pdf (accessed on 11 July 2017).

73. Zheng, K.; Leung, V.C.M.; Yang, L.L.; Chatzimisios, P. Guest Editorial Special Issue on 5G Wireless Systems with Massive MIMO. *IEEE Syst. J.* **2017**, *11*. [CrossRef]

74. Ge, X.; Cheng, H.; Guizani, M.; Han, T. 5G Wireless Backhaul Networks: Challenges and Research Advances. *IEEE Netw.* **2014**, *28*, 6–11. [CrossRef]

75. Liu, J.; Kato, N.; Ma, J.; Kadowaki, N. Device-to-Device Communication in LTE-Advanced Networks: A Survey. *IEEE Commun. Surv. Tutor.* **2014**, *17*, 1923–1940. [CrossRef]

76. Pérez-Romero, J.; Sallent, O.; Agusti, R. Enhancing Cellular Coverage through Opportunistic Networks with Learning Mechanisms. In Proceedings of the IEEE Global Telecommunication Conference, Atlanta, GA, USA, 9–13 December 2013.

77. Asadi, A.; Wang, Q.; Mancuso, V. A Survey on Device-to-Device Communication in Cellular Networks. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1801–1819. [CrossRef]

78. Chen, T.; Kunnari, E.; Tapani, R. Device-to-Device Communication for LTE-Advanced Network System. Available online: https://tel.archives-ouvertes.fr/tel-00983507/document (accessed on 10 July 2017).

79. Hicham, M.; Abghour, N.; Ouzzif, M. Device-to-Device (D2D) Communication under LTE-Advanced Networks. *Int. J. Wirel. Mob. Netw.* **2016**, *8*, 11–22. [CrossRef]

80. Noura, M.; Nordin, R. A Survey on Interface for Device-to-Device (D2D) Communication and its Challenges in 5G Networks. *J. Netw. Comput. Appl.* **2016**. [CrossRef]

81. Mozaffari, M.; Saad, W.; Bennis, M.; Debbah, M. Mobile Unmanned Aerial Vehicles (UAVs) for Energy-Efficient Internet of Things Communications. Unpublished work, 2017.

82. Hayat, S.; Yanmaz, E.; Muzaffar, R. Survey on Unmanned Aerial Vehicle Networks for Civil Applications: A Communications Viewpoint. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2624–2661. [CrossRef]

83. Larsson, E.G.; Edfors, O.; Tufvesson, F.; Marzetta, T.L. A Massive MIMO for Next Generation Wireless Systems. *IEEE Commun. Mag.* **2016**, *52*, 186–195. [CrossRef]

84. Jovanovic, A.; Li, J.; Richardson, T. Visible light Communication: Opportunities, Challenges and the Path to Market. *IEEE Commun. Mag.* **2013**, *51*, 130–135. [CrossRef]

85. Popovicki, G.; Lauridsen, M.; Soret, B.; Pedersen, K.I.; Mogensen, P. Automation for on-Road Vehicles: Use Cases and Requirements for Radio Design. In Proceedings of the IEEE 82nd Vehicular Technology Conference (VTC Fall), Boston, MA, USA, 6–9 September 2015.

86. Abboud, K.; Omar, H.A.; Zhuang, W. Interworking of DSRC and Cellular Network Technologies for V2X Communications: A Survey. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9457–9470. [CrossRef]

87. Zheng, K.; Zhao, L.; Mei, J.; Dohler, M.; Xiang, W.; Peng, Y. 10 Gb/s HetSNets with Millimeter-Wave Communications: Access and Networking-Challenges and Protocols. *IEEE Commun. Mag.* **2015**, *53*, 222–231. [CrossRef]

88. Jaber, M.; Imran, M.A.; Tafazolli, R.; Tukmanov, A. 5G Backhaul Challendes and Emerging Research Direction: A Survey. *IEEE Access* **2014**, *4*, 1743–1766. [CrossRef]

89. Abdul Salam, S.; Mahmud, S.A.; Khan, G.M.; Raweshidy, H.S. M2M Communication in Smart Grids: Implementation Scenarios and Performance Analysis. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Paris, France, 1 April 2012.

90. 3GPP. Technical Specification Group Services and System Aspects (Release 8). Available online: http://www.arib.or.jp/IMT-2000/V720Mar09/5_Appendix/Rel8/23/23521-800.pdf (accessed on 10 July 2017).

91. Ratasuk, R.; Tan, J.; Ghosh, A. Coverage and capacity analysis for machine type communications in LTE. In Proceedings of the IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, Japan, 6–9 May 2012.

92. Wu, G.; Talwar, S.; Johnsson, K.; Himayat, N.; Johnson, K.D. M2M: From Mobile to Embedded Internet. *IEEE Commun. Mag.* **2011**, *49*, 36–43. [CrossRef]

93. Chen, M.; Wan, J.; Li, F. Machine-To-Machine Communications: Architectures, Standards and Applications. *KSII Trans. Internet Inf. Syst.* **2012**, *6*, 480–497. [CrossRef]

94. Marjanovic, M.; Skorin-Kapov, L.; Pripužić, K.; Antonić, A.; Podnar Žarko, I. Energy-Aware and Quality-Driven Sensor Management For Green Mobile Crowd Sensing. *J. Netw. Comput. Appl. Arch.* **2015**, *59*, 96–108. [CrossRef]

95. 5GPPP. View on 5G Architecture. Available online: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-For-public-consultation.pdf (accessed on 10 July 2017).

96. 5GPP. 5G Empowering Vertical Industries. Available online: https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf (accessed on 10 July 2017).

97. 4G Americas. Cellular Technologies Enabling the Internet of Things. Available online: http://www.5gamericas.org/files/6014/4683/4670/4G_Americas_Cellular_Technologies_Enabling_the_IoT_White_Paper_-_November_2015.pdf (accessed on 10 July 2017).

98. Carlesso, M.; Antonopoilos, A.; Granelli, F.; Verikoukis, C. Uplink Scheduling form Smart Metering and Real-Time Traffic Coexistence in LTE Networks. In Proceedings of the International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 820–825.

99. Islam, R.; Avazov, N.; Dobre, O.A.; Kwak, K.S. Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges. *IEEE Commun. Surv. Tutor.* **2016**, *19*, 721–742. [CrossRef]

100. Liu, Y.; Zhang, Y.; Yu, R.; Xie, R. Integrated Energy and Spectrum Harvesting for 5G Wireless Communications. *IEEE Netw.* **2015**, *29*, 75–81. [CrossRef]

101. Kong, L.; Khan, M.K.; Wu, F.; Chen, G.; Zeng, P. Millimeter-Wave Wireless Communications for IoT-Cloud Supported Autonomous Vehicles: Overview, Design, and Challenges. *IEEE Commun. Mag.* **2017**, *55*, 62–68. [CrossRef]

102. Costa-Perez, X.; Garcia-Saavedra, A.; Li, X.; de la Oliva, A.; Iovanna, P.; Deiß, T.; di Giglio, A.; Mourad, A. 5G-Crosshaul: An SDN/NFV Integrated Fronthaul/Backhaul Transport Network Architecture. *IEEE Wirel. Commun.* **2017**, *24*, 38–45. [CrossRef]

103. Trivisonno, R.; Guerzoni, R.; Vaishnavi, I.; Soldani, D. Towards Zero Latency Software Defined 5G Networks. In Proceedings of the IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015.

104. Esaki, H.; Nakamura, R. Overlaying and Slicing for Iot Era Based on Internet's End-To-End Discipline. In Proceedings of the IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), Osaka, Japan, 12–14 June 2017.

105. 5GPP. 5G and Media & Entertainment. Available online: https://5g-ppp.eu/wp-content/uploads/2016/02/5G-PPP-White-Paper-on-Media-Entertainment-Vertical-Sector.pdf (accessed on 10 July 2017).

106. GSA. The Road to 5G: Frivers, Applications, Requirements and Technical Development. Available online: http://www.huawei.com/minisite/5g/img/GSA_the_Road_to_5G.pdf (accessed on 10 July 2017).

107. Erman, B.; Yiu, S. Modeling 5G Wireless Network Service Reliability Prediction with Bayesian Network. In Proceedings of the 2016 IEEE International Workshop Technical Committee on Communications Quality and Reliability, Stevenson, WA, USA, 10–12 May 2016.

108. NOKIA. 5G Use Cases and Requirements White Paper. Available online: https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf (accessed on 10 July 2017).

109. Motlagh, N.H.; Bagaa, M.; Taleb, T.; Song, J. Connection Steering Mechanism between Mobile Networks for Reliable UAV's Iot Platform. In Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, 21–25 July 2017.

110. Kaltenberger, F.; Knopp, R.; Vitiello, C.; Danneberg, M.; Festag, A. Experimental Anaysis of 5G Candidate Waveform and Their Coexistence with 4G Systems. In Proceedings of the Joint NEWCOM/COST Workshop on Wireless Communications JNCW, Barcelona, Spain, 14–15 October 2016.

111. Serra, J.; Pubill, D.; Antonopoulos, A.; Verikoukis, C. Smart HVAC Control in IoT: Energy Consumption Minimization with User Comfort Constraints. *Sci. World J.* **2014**, *2014*. [CrossRef] [PubMed]

112. Nikolich, P.; Lin, C.; Korhonen, J.; Marks, R.; Tye, B.; Li, G.; Ni, J.; Zhang, S. Standards for 5G and Beyond: Their Use Cases and Applications. *IEEE 5G Tech Focus* **2017**, *1*. Available online: https://5g.ieee.org/tech-focus/june-2017/standards-for-5g-and-beyond (accessed on 9 September 2017).

113. Kusume, K.; Fallgren, M.; Queseth, O.; Braun, V.; Gozalvez-Serrano, D.; Korthals, I.; Zimmermann, G.; Schubert, M.; Hossain, M.I.; Widaa, A.A.; et al. Updated Scenarios, Requirements and KPIs for 5G Mobile and Wireless System with Recommendations for Future Investigations. METIS_D1.5_v1. METIS. Available online: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.5_v1.pdf (accessed on 10 July 2017).

114. 5GPP. 5G Vision and Requirements. Available online: https://5g-ppp.eu/imt-2020-5g-promotion-group-and-5g-ppp-announce-memorandum-of-understanding-for-5g/ (accessed on 10 July 2017).

115. Yang, C.; Li, J. *Interference Mitigation and Energy Management in 5G Heterogeneous Cellular Networks*; ICI Global: Harshey, PA, USA, 2017; pp. 1–362.

116. Ericsson. 5G Energy Performance. Available online: http://www.5gamericas.org/files/3114/3898/7332/Ericsson_wp-5g-energy-performance.pdf (accessed on 10 July 2017).

117. Shen, K.; Yu, W. Distributed Pricing-Based User Association for Downlink Hterogeneous Cellular Networks. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1100–1113. [CrossRef]

118. Nguyen, H.V.; Le, L.B. Distributed Base Station Association and Power Control for Heterogeneous Cellular Networks. *IEEE Trans. Veh. Technol.* **2014**, *63*, 282–296. [CrossRef]

119. Singh, S.; Andrews, J.G. Joint Resource Portitioning and Offloading in Heterogeneous Cellular Networks. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 888–901. [CrossRef]

120. Elsawy, H.; Hossain, E.; Kim, D.I. HetNets with Cognitive Small Cells: User Offloading and Distributed Channel Access Technique. *IEEE Commun. Mag.* **2013**, *51*, 28–36. [CrossRef]

121. Ji, Z.; Li, K. Cognitive Radios for Dynamic Spectrum Access-Dynamic Spectrum Sharing: A Game Theoretical Overview. *IEEE Commun. Mag.* **2007**, *45*, 88–94. [CrossRef]

122. Aijazand, A.; Aghvami, A.H. PRMA-Based Cognitive Machine-to-Machine Communications in Smart Grid Networks. *IEEE Trans. Veh. Technol.* **2015**, *64*, 3608–3623. [CrossRef]

123. Okasaka, S.; Weiler, R.J.; Keusgen, W.; Pudeyev, A.; Maltsev, A.; Karls, I.; Sakaguchi, K. Proof-of-Concept of a Millimeter-Wave Integrated Heterogeneous Network for 5G Cellular. *Sensors* **2016**, *9*. [CrossRef] [PubMed]

124. Qi, S.; Tesanovic, M. FQAM-FBMC Design and Its Application to Machine Type Communication. In Proceedings of the IEEE 27th Annual International Symposium on PIMRC, Valencia, Spain, 4–8 September 2016.

125. Domouchtsidis, S.G.; Ntouni, G.D.; Kapinas, V.M.; Karagiannidis, G.K. OFDM-IM vs. FQAM: A Comparative Analysis. In Proceedings of the International Conference on Telecommunications, Thessaloniki, Greece, 16–18 May 2016.

126. Hong, S.; Sagong, M.; Lim, C.V.; Cheun, K.; Cho, S. FQAM: A Modulation Scheme for Beyond 4G Cellular Wireless Communication Systems. In Proceedings of the IEEE GC Wkshps, Atlanta, GA, USA, 9–13 December 2013.

127. Perlman, S.; Forenza, A. pCell Wireless Reinvented. Arthemis Networks. Available online: http://www.rearden.com/DIDO/DIDO_White_Paper_110727.pdf (accessed on 10 July 2017).

128. Mesleh, R.Y.; Hass, H.; Sinanovic, S.; Ahn, C.W.; Yun, S. Spatial Modulation. *IEEE Trans. Veh. Technol.* **2008**, *57*, 2228–2241. [CrossRef]

129. Maunder, R.G. The 5G Channel Code Contenders. Available online: https://www.accelercomm.com/sites/accelercomm.com/files/WhitePaper09Aug2016.pdf (accessed on 10 July 2017).

130. Scholl, S.; Weithoffer, S.; When, N. Advanced Iterative Channel Coding Schemes: When Shannon meets Moore. In Proceedings of the 9th International Symposium on Turbo Codes & Iterative Information, Brest, France, 5–9 September 2016.

131. Korpi, D.; Tamminen, J.; Tarunen, M.; Huusari, T.; Choi, Y.-S.; Anttila, L.; Talwar, S.; Valkama, M. Full-Duplex Mobile Device: Pushing the Limit's. *IEEE Commun. Mag.* **2016**, *54*, 80–87. [CrossRef]

132. Hong, S.S.; Brand, J.; Choi, J.I.; Jain, M.; Mehl, J. Applications of Self-Interference Cancellation in 5G and Beyond. *IEEE Commun. Mag.* **2014**, *52*, 114–121. [CrossRef]

133. Duarte, M.; Dick, C.; Sabharwal, A. Experiment-Driven Characterization of Full-Duplex Wireless Systems. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 4296–4307. [CrossRef]

134. Bharadia, D.; McMilin, E.; Katti, S. Full Duplex Radios. In Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM, Hong Kong, China, 12–16 August 2013.

135. Čelik, H.; Sung, K.W. On the Feasibility of Blind Dynamic TDD in Ultra-Dense Wireless Networks. In Proceedings of the 2015 IEEE 81st VehicularTechnology Conference (VTC Spring), Glasgow, UK, 11–14 May 2015.

136. Osti, P.; Aalto, S.; Lassila, P. Flow-Level Modeling and Optimization of Intercell Coordination with Dynamic TDD. In Proceedings of the 10th ACM symposium on QoS and security for wireless and mobile networks Q2SWinet '14, Montreal, QC, Canada, 21–26 September 2014.

137. Kerttula, J.; Marttinen, A.; Ruttik, K.; Jäntti, R.; Alam, N. Dynamic TDD on LTE Small Cells. *EURASIP J. Wirel. Commun. Netw.* **2016**, *194*. [CrossRef]

138. Gupta, A.K.; Kulkarni, M.N.; Visotsky, E.; Vook, F.W.; Ghosh, A.; Andrews, J.G.; Heath, R.W. Rate Analysis and Feasibility of Dynamic TDD in 5G Cellular Systems. In Proceedings of the 2016 IEEE International Conference on Communications, Cuala Lumpur, Malaysia, 22–27 May 2016.

139. Sun, H.; Wildemeersch, M.; Sheng, M.; Quek, T.K. D2D Enchanced Heterogeneus Cellular Networks with Dynamic TDD. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 4204–4218. [CrossRef]

140. Sun, H.; Wildemeersch, M.; Sheng, M.; Quek, T.K.; Li, J. Traffic Adaptation and Energy Efficiency for Small Cell Networks with Dynamic TDD. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 3234–3251. [CrossRef]

141. Čelik, H.; Song, K.W. Joint Transmission with Dummy Symbols for Dynamic TDD in Ultra-Dense Deployment. Available online: https://aps.arxiv.org/abs/1702.06045 (accessed on 10 July 2017).

142. Li, J.; Farahvash, S.; Kavehrad, M.; Valenzuel, R. Dynamic TDD and Fixed Cellular Networks. *IEEE Commun. Lett.* **2002**, *4*, 218–220. [CrossRef]

143. Shen, Z.; Khoryae, A.; Eriksson, E.; Pan, X. Dynamic Uplink-Downlink Configuration and Interference Management in TD-LTE. *IEEE Commun. Mag.* **2012**, *50*, 51–59. [CrossRef]

144. Chunlin, Y.; Zhifeng, Y.; Weimin, L.; Yifei, Y. Non-Orthogonal Multiple Access Schemes for 5G. *ZTE Commun.* **2016**, *14*, 11–16. [CrossRef]

145. 3GPP. Discussion on Multiple Access for Radio Interface. Available online: http://www.3gpp.org/DynaReport/TDocExMtg--R1-84b--31661.htm (accessed on 10 July 2017).

146. Yuan, Z.; Yu, G.; Li, W.; Yuan, Y.; Wang, X.; Xu, J. Multi-user Shared Access for Internet of Things. In Proceedings of the 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China, 15–18 May 2016.

147. 3GPP. Receiver Implementation for MUSA. 3GPP. Available online: http://www.3gpp.org/DynaReport/TDocExMtg--R1-85--31662.htm (accessed on 10 July 2017).

148. 3GPP. Contetion-Based Non-Orthogonal Multiple Access for UL mMTC. Available online: http://www.3gpp.org/DynaReport/TDocExMtg--R1-85--31662.htm (accessed on 10 July 2017).

149. 3GPP. Resource Spread Multiple Access. Available online: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_85/Docs/ (accessed on 10 July 2017).

150. Taherzadeh, M.; Nikopour, H.; Bayesteh, A.; Baligh, H. SCMA Codebook Design. In Proceedings of the IEEE International Symposium on Personal, Indoor And Mobile Radio Communications, London, UK, 14–17 September 2014.

151. Nikopour, H.; Baligh, H. Sparse Code Multiple Acccess. In Proceedings of the 2014 IEEE 80th Vehicular Technology Conference, Vancouver, BC, Canada, 14–17 September 2014.

152. Future Mobile Communication Forum. 5G Whitepaper v2.0, Part d-Alternative Multiple Access v1. Available online: http://www.future-forum.org/dl/151106/whitepaper.rar (accessed on 10 July 2017).

153. 3GPP. Candidate Solution for New Multiple Access. 3GPP. Available online: http://portal.3gpp.org/ngppapp/CreateTdoc.aspx?mode=view&contributionId=697587 (accessed on 10 July 2017).

154. Dai, X.; Chen, S.; Sun, S.; Kang, S.; Wang, Y.; Shen, Z.; Xu, Z. Successive Interference Cancelation Amenable Multiple Access (SAMA) for Future Wireless Communications. In Proceedings of the IEEE International Conference on Communication Systems, Macau, China, 19–21 November 2014.

155. Dai, X. Successive Interference Cancellation Amenable Space-Time Codes with Good Multiplexing-Diversity Tradeoff. *Wirel. Pers. Commun.* **2010**, *55*, 645–654. [CrossRef]

156. Li, P.; Liu, L.; Wu, K.; Leung, W.K. On Interleave-Division Multiple-Access. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017.

157. Li, P.; Liu, L.; Wu, K.; Leung, W.K. Interleave Division Multiple-Access. *IEEE Trans. Wirel. Commun.* **2006**, *5*, 938–947. [CrossRef]

158. Rabee, F.A.; Dayaslioglu, K.; Gitlin, R. The Optimum Received Power Levels of Uplink Non-Orthogonal Multiple Access (NOMA) Signals. In Proceedings of the 2017 IEEE 18th Wireless and Microwave Technology Conference (WAMICON), Cocoa Beach, FL, USA, 24–25 April 2017.

159. Saito, Y.; Kishiyama, Y.; Benjebbour, A.; Nakamura, T.; Li, A.; Higuchi, K. Non-orthogonal Multiple Access (NOMA) for Cellular Future Radio Access. In Proceedings of the 77th IEEE Vehicular Technology Conference, Dresden, Germany, 2–5 June 2013.

160. Choi, J. On Multiple Access Using H-ARQ with SIC Techniques for Wireless Ad Hoc Networks. *Wirel. Pers. Commun.* **2013**, *69*, 187–212. [CrossRef]

161. Choi, J. Non-orthogonal Multiple Access in Downlink Coordinated Two-point Systems. *IEEE Commun. Lett.* **2014**, *18*, 313–316. [CrossRef]

162. Sedtheetorn, P.; Chulajata, T. Accurate Spectral Efficiency Analysis for Non Orthogonal Multiple Access. *Trans. Adv. Commun. Technol.* **2016**, *5*, 844–850.

163. Wild, T.; Schaich, F.; Chen, Y. 5G Air Interface Design Based on Universal Filtered (UF-) OFDM. In Proceedings of the 19th International Conference in DSP, Hong Kong, China, 20–23 August 2014.

164. Ding, Z.; Yang, Z.; Fan, P.; Poor, H.V. On the Performance of Non-orthogonal Multiple Access in 5G Systems with Randomly Deployed Users. *IEEE Signal Process. Lett.* **2014**, *21*, 1501–1505. [CrossRef]

165. Zhao, Z.; Schellmann, M.; Wang, Q.; Gong, X.; Boehnke, R.; Xu, W. Pulse shaped OFDM for Asynchronous Uplink Access. In Proceedings of the Asilomar Conference Signals, Systems and Computers, Pacific Grove, CA, USA, 8–11 November 2015.

166. Cover, T.M.; El Gamel, A. Capacity Theorems for The Relay Channel. *IEEE Trans. Inf. Theory* **1979**, *25*, 572–584. [CrossRef]

167. Cover, T.M.; Leung, C.S.K. An Achievable Rate Region for the Multiple-access Channel with Feedback. *IEEE Trans. Inf. Theory* **1981**, *27*, 292–298. [CrossRef]

168. Ahlswede, R. Multiway Communication Channels. In Proceedings of the 2nd International Symposium on Information Theory, Tsahkadsor, Armenian SSR, 2–8 September 1971.

169. Xie, L.L.; Kumar, P.R. An Achievable Rate for the Multiple-level Relay Channel. *IEEE Trans. Inf. Theory* **2005**, *51*, 1348–1358. [CrossRef]

170. Kim, K.T.; Ahn, S.-K.; Kim, Y.-H.; Park, H.; Wang, L.; Chen, C.-Y.; Park, J. Adaptive Sliding-Window Coded Modulation in Cellular Networks. In Proceedings of the IEEE GLOBECOM, San Diego, CA, USA, 6–10 December 2015.

171. Thurfjell, M.; Ericsson, M.; de Bruin, P. Network Densification Impact on System Capacity. In Proceedings of the 2015 IEEE 81st Vehicular Technology Conference, Glasgow, Scotland, 11–14 May 2015.

172. Darwish, A.; Ibrahim, A.S. Capacity Improvements via Indoor Small Cells. In Proceedings of the 2014 International Wireless Communications and Mobile Computing Conference (IWCMC), Nicosia, Cyprus, 4–8 August 2014.

173. Hoadley, J.; Maveddat, P. Enabling Small Cell Deployment with Hetnet. *IEEE Wirel. Commun.* **2012**, *19*, 4–5. [CrossRef]

174. Hughes, M.; Jovanovic, V.M. Small Cells-Effective Capacity Relief Option for Heterogeneous Networks. In Proceedings of the 2012 IEEE Vehicular Technology Conference, Quebec City, QC, Canada, 3–6 September 2012.

175. Sui, Y.; Guvenc, I.; Svensson, T. Interference Management for Moving Networks in Ultra-Dense Urban Scenarios. *EURASIP J. Wirel. Commun. Netw.* **2015**, *111*. [CrossRef]

176. Sui, Y.; Vihrälä, J.; Papadogiannis, A.; Sternad, M.; Svensson, T. Moving Cells: A Promising Solution to Boost Performance for Vehicular Users. *IEEE Commun. Mag.* **2013**, *51*, 62–68. [CrossRef]

177. Sambo, Y.A.; Shakir, M.Z.; Héliot, F.; Imran, M.A.; Mumtaz, S.; Qaraqe, K.A. Device-to-Device Communication in Heterogeneous Networks. In *Smart Device to Smart Device Communication*; Mumtaz, S., Rodriguez, J., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 219–235.

178. Liu, X.; Andrews, J.G.; Ghosh, A. Spectrum Sharing for Device- to-Device Communication in Cellular Networks. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 6272–6740. [CrossRef]

179. Zafar, B.; Gherekhloo, S.; Haardt, M. Analysis of Multihop Relaying Networks: Communication Between Range-Limited and Cooperative Nodes. *IEEE Veh. Technol. Mag.* **2012**, *7*, 40–47. [CrossRef]

180. Kamalinejad, P.; Mahapatra, C.; Sheng, Z.; Mirabbasi, S.; Leung, V.C.M.; Guan, Y.L. Wireless Energy Harvesting for the Internet of Things. *IEEE Commun. Mag.* **2015**, *53*, 102–108. [CrossRef]

181. Mekikis, P.-V.; Antonopoulos, A.; Kartsakli, E.; Lalos, A.S.; Alonso, L.; Verikoukis, C. Information Exchange in Randomly Deploys Dense WSNs with Wireless Energy Harvesting Capabilities. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 3008–3018. [CrossRef]

182. Mekikis, P.-V.; Antonopoulos, A.; Alonso, L.; Verikoukis, C. Stochastic Modeling of Wireless Charged Wearables for Reliable Health Monitoring In Hospital Environments. In Proceedings of the International Conference on Communications (ICC), Paris, France, 21–25 May 2017.

183. Li, Y.; Cai, L. UAV-Assisted Dynamic Coverage in a Heterogeneous Cellular System. *IEEE Netw.* **2017**, *31*, 56–61. [CrossRef]

184. Messous, M.-A.; Senouci, S.-M.; Sedjelmaci, H. Network Connectivity and Area Coverage for UAV Fleet Mobility Model with Energy Constraint. In Proceedings of the Wireless Communications and Networking Conference (WCNC), Doha, Qatar, 3–6 April 2016.

185. Mekikis, P.-V.; Kartsakli, E.; Alonso, L.; Verikoukis, C. Flexible Aerial Relay Nodes for Communication Recovery and D2D Relaying. In Proceedings of the IEEE 5th Global Conference on Consumer Electronics, Kyoto, Japan, 11–14 October 2016.

186. Zeng, Y.; Zhang, R.; Lim, T.J. Wireless Communications with Unmanned Aerial Vehicles: Opportunities and Challenges. *IEEE Commun. Mag.* **2016**, *54*, 36–42. [CrossRef]

187. Yang, X.; Lu, W.-J.; Wang, N.; Nieman, K.; Jin, S.; Zhu, H.; Mu, X.; Wong, I.; Huang, Y.; You, X. Design and Implementation of a TDD-Based 128-Antenna Massive MIMO Prototyping System. Available online: http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1608.07362 (accessed on 11 July 2017).

188. Chen, B.; Zhu, C.; Shu, L.; Su, M.; Wei, J.; Leung, V.C.M.; Rodriques, J.J.P.C. Uplink Transmission for Lightweight Single-antenna UEs in the Presence of a Massive MIMO eavesdropper. *IEEE Access* **2016**, *99*, 2169–3536. [CrossRef]

189. Wu, S.; Wang, H.; Youn, C.H. Visible Light Communications for 5G Wireless Networking Systems: From Fixed to Mobile. *IEEE Access* **2014**, *28*, 41–45. [CrossRef]

190. Feng, L.; Hu, R.Q.; Wang, J.; Xu, P.; Qian, Y. Applying VLC in 5G Networks: Architectures and Key Technologies. *IEEE Access* **2016**, *30*, 77–83. [CrossRef]

191. Grobe, L.; Paraskevopoulos, A.; Hilt, J.; Schulz, D.; Lassak, F.; Hartlieb, F.; Kottke, C.; Jungnickel, V.; Langer, K.-D. High-Speed Visible Light Communication Systems. *IEEE Commun. Mag.* **2013**, *51*, 60–66. [CrossRef]

192. Li, X.; Zhang, R.; Wang, J.; Hanzo, L. Cell-Centric and User-Centric Multi-User Scheduling in Visible Light Communication Aided Networks. In Proceedings of the IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015.

193. Katriniok, A.; Kleibaum, P.; Res, C.; Eckstein, L. Automation of Road Vehicles Using V2X: An Application to Intersection Automation. Available online: http://papers.sae.org/2017-01-0078/ (accessed on 10 July 2017).

194. Ghods, A.; Severi, S.; Abreu, G. Localization in V2X Communication Networks. In Proceedings of the IEEE Intelligent Vehicles Symposium, Gothenburg, Sweden, 19–22 June 2016.

195. Goebel, N.; Bialon, R.; Mauve, M.; Graffi, K. Coupled Simulation of Mobile Cellular Networks, Road Traffic and V2X Applications Using Traces. In Proceedings of the International Conference on Communications, Kuala Lumpur, Malaysia, 22–27 May 2016.

196. Xiao, Z.; Liu, H.; Havyarimana, V.; Li, T.; Wang, D. Analytical Study on Multi-Tier 5G Heterogeneous Small Cell Networks: Coverage Performance and Energy Efficiency. *Sensors* **2016**, *11*. [CrossRef] [PubMed]

197. Choi, J.I.; Jain, M.; Srinivasan, K.; Levis, P.; Katti, S. Achieving Single Channel, Full Duplex Wireless Communication. In Proceedings of the sixteenth annual international conference on Mobile computing and networking, Chicago, IL, USA, 20–24 September 2010.

198. Song, Q.; Nuaymi, L.; Lagrange, X. Survey of Radio Resource Management Issues and Proposals for Energy-Efficient Cellular Network That Will Cover Billions of Machines. *Wirel. Commun. Netw.* **2016**, *140*. [CrossRef]

199. Costa, G.; Miao, G. Context-Aware Machine-To-Machine Communications. In Proceedings of the 33rd Annual IEEE International Conference on Computer Communications (INFOCOM WKSHPS), Toronto, ON, Canada, 27 April–2 May 2014.

200. Agyapong, P.K.; Iwamura, M.; Staehle, D.; Kiess, W.; Benjebbour, A. Design Considerations for a 5G Network Architecture. *IEEE Commun. Mag.* **2014**, *52*, 65–75. [CrossRef]

201. Vermesan, O.; Friass, P.; Guillemin, P.; Giaffreda, R.; Grindvoll, H.; Eisenhauer, M.; Serrano, M.; Moessner, K.; Spirito, M.; Blystad, L.-C.; et al. *Building the Hyperconnected Society—IoT Research and Innovation Value Chains, Ecosystems and Markets*; Vermesan, O., Friass, P., Eds.; River Publishers: Aalborg, Denmark, 2015; Volume 43.

202. Panwar, N.; Sharma, S.; Singh, A.K. A Survey on 5G: The Next Generation of Mobile Communication. *Elsevier Phys. Commun.* **2016**, *18*, 64–84. [CrossRef]

203. Thuemmler, C.; Gavrasm, A.; Jumelle, A.K.L.; Paulin, A.; Sadique, A.; Schneider, A.; Fedell, C.; Abraham, D.; Trossen, D. 5G and e-Health. Available online: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-eHealth-Vertical-Sector.pdf (accessed on 10 July 2017).

204. Rose, K.; Eldridge, S.; Chaplin, L. The Internet of Things: An Overview. Available online: https://www.internetsociety.org/sites/default/files/ISOC-IoT-Overview-20151014_0.pdf (accessed on 10 July 2017).

205. Huawei. 5G Opening up New Business Opportunities. Available online: http://www-file.huawei.com/-/media/CORPORATE/PDF/x-lab/10-5g-opening-up-new-business-opportunities-en.pdf?la=en (accessed on 10 July 2017).

206. Lin, H.; Shao, J.; Zhang, C.; Fang, Y. CAM: Cloud-Assisted Privacy Preserving Mobile Health Monitoring. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 985–997. [CrossRef]

207. NGMN Alliance. NGMN 5G White Paper. Available online: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf (accessed on 10 July 2017).

208. Le, N.T.; Hossain, M.A.; Islam, A.; Kim, D.-Y.; Choi, Y.-J.; Jang, Y.M. Survey of Promising Technologies for 5G Networks. *Hindawi* **2016**, *2016*. [CrossRef]

209. Militano, L.; Araniti, G.; Condoluci, M.; Farris, I.; Iera, A. Device-to-Device Communications for 5G Internet of Things. *EAI Endorsed Trans. Internet Things* **2015**, *1*. [CrossRef]

210. Shariatmadari, H.; Ratasuk, R.; Iraji, S.; Laya, A.; Taleb, T.; Ghosh, A. Machine-type Communications: Current Status and Future Perspective toward 5G Systems. *IEEE Commun. Mag.* **2015**, *53*, 10–17. [CrossRef]

211. 5GPPP. 5G and the Factories of the Future. Available online: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf (accessed on 10 July 2017).

212. Elayoubi, S.E.; Fallgren, M.; Spapis, P.; Zimmermann, G.; Martín-Sacristán, D.; Yang, C.; Jeux, S.; Agyapong, P.; Campoy, L.; Qi, Y.; et al. 5G Service Requirements and Operational Use Cases: Analysis and METIS II Vision. In Proceedings of the 2016 European Conference on Networks and Communications (EuCNC), Athens, Greece, 27–30 June 2016.

213. Blanco, B.; Fajardo, H.O.; Giannis, I.; Kafetzakis, E.; Pneg, S.; Perez-Romero, J.; Trajkovska, I.; Khodashenas, P.S.; Goratti, L.; Paolino, M.; et al. Technology Pillars in the Architecture of Future 5G Mobile Networks: NFV, MEC and SDN. *Comput. Stand. Interfaces* **2017**, *54*, 216–228. [CrossRef]

214. ITU-R. Recommendation ITU-R M.2083-0. IMT Vision: Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond. Available online: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf (accessed on 10 July 2017).

215. Le, L.B.; Lau, V.; Jorswieck, E.; Dao, N.-D.; Haghighat, A.; Kim, D.I.; Le-Ngoc, T. Enabling 5G Mobile Wireless Technologies. *EURASIP J. Wirel. Commun. Netw.* **2015**, *218*. [CrossRef]

216. Liu, G.; Jiang, D. 5G: Vision and Requirements for Mobile Communication System towards Year 2020. *Chin. J. Eng.* **2016**, *2016*, 5974586. [CrossRef]

217. Bogale, T.E.; Le, L.B. Massive MIMO and mmWave for 5G Wireless HetNet: Potential Benefit's and Challenges. *IEEE Veh. Technol. Mag.* **2016**, *11*, 64–75. [CrossRef]

218. Banelli, P.; Buzzi, S.; Colavolpe, G.; Modenini, A.; Rusek, F.; Ugolini, A. Modulation Formats and Waveforms for 5G Networks: Who Will Be the Heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency. *IEEE Signal Process. Mag.* **2014**, *31*, 80–93. [CrossRef]

219. Palattella, M.R.; Dohler, M.; Grieco, A.; Rizzo, G.; Torsner, J.; Engel, T.; Ladidm, L. Internet of Things in the 5G Era: Enablers, Architecture and Business Models. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 510–527. [CrossRef]

220. Pries, R.; Morper, H.-J.; Galambosi, N.; Jarschel, M. Network as a Service—A Demo on 5G Network Slicing. In Proceedings of the 28th International Teletraffic Congress (ITC 28), Würzburg, Germany, 12–16 September 2016.

221. Caraguay, A.L.V.; Peral, A.B.; Lopez, L.I.B. SDN: Evolution and Opportunities in the Development IoT Applications. *Int. J. Distrib. Sens. Netw.* **2014**, *10*. [CrossRef]

222. Bizanis, N.; Kuipers, F.A. SDN and Virtualization Solutions for the Internet of Things: A Survey. *IEEE Access* **2016**, *4*, 5591–5606. [CrossRef]

223. Stancu, A.; Vulpe, A.; Fratu, O.; Halunga, S. Default Values Mediator Used for a Wireless Transport SDN Proof of Concept. In Proceedings of the IEEE Conference on Standards for Communications and Networking (CSCN), Berlin, Germany, 31 October–2 November 2016.

224. Warren, D.; Dewar, C. Understanding 5G: Perspectives on Future Technological Advancements in Mobile. Available online: https://www.gsmaintelligence.com/research/?file=141208-5g.pdf&download (accessed on 10 July 2017).

225. Rappaport, T.S.; Heath, R.V.; Daniels, R.C.; Murdock, J.N. *Millimeter Wave Wireless Communications*, 1st ed.; Pearson: Harshey, PA, USA, 2017; pp. 1–657.

226. Thakare, S. Research Challenges of Millimeter Wave Technology in 5G Cellular System. In Proceedings of the 4th International Conference on Recent Trends in Engineering & Technology (ICRTET), Nashik, India, 2–4 July 2015.

227. Ye, J.; He, J.; Ge, X.; Chen, M. Energy Efficiency Analysis of 5G Ultra-dense Networks Based on Random Way Point Mobility Models. In Proceedings of the 2016 19th International Symposium on Wireless Personal Multimedia Communications, Shenzhen, China, 14–16 November 2016.

228. Witrisal, K.; Meissner, P.; Leitinger, E.; Shen, Y.; Gustafson, C.; Tufvesson, F.; Haneda, K.; Dardari, D.; Molisch, A.F.; Conti, A.; et al. High-Accurany Localization for Assisted Living: 5G Systems Will Turn Multipath Channels from Foe to Friend. *IEEE Signal Process. Mag.* **2016**, *33*, 59–70. [CrossRef]

229. 5GPP. 5G Automotive Vision. Available online: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf (accessed on 10 July 2017).

230. Siaud, I.; Umner-Moll, A.M. Green-Oriented Multi-Techno Link Adaptation Metrics for 5G Heterogeneous Networks. *EURASIP J. Wirel. Commun. Netw.* **2016**, *92*. [CrossRef]

231. Le, L.B.; Niyato, D.; Hossain, E.; Kim, D.I.; Hoang, D.T. QoS-Aware and Energy-Efficient Resource Management in OFDMA Femtocells. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 180–194. [CrossRef]

232. Ye, Q.; Rong, B.; Chen, Y.; Al-Shalash, M.; Caramanis, C.; Andrews, J.G. User Association for Load Balancing in Heterogeneous Cellular Networks. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 2706–2716. [CrossRef]

233. Mavromoustakis, C.X.; Bourdena, A.; Mastorakis, G.; Pallis, E.; Kormentzas, G. Coverage An Energy-Aware Scheme for Efficient Spectrum Utilization in A 5G Mobile Cognitive Radio Network Architecture. *Telecommun. Syst.* **2015**, *59*, 63–75. [CrossRef]

234. Wu, S.; Wang, Y.; Al Imari, M.; Nekovee, M. FQAM: Frequency and Quadrature Amplitude Modulation for 5G Networks. In Proceedings of the 2016 European Conference on Networks and Communications (EuCNC), Athens, Greece, 27–30 June 2016.

235. Al-Imari, M.; Mouhouche, B.; Al.Imari, M.; Nekovee, M.F. Non-Orthogonal FQAM for Multiple Access in the Uplink of 5G Wireless Networks. In Proceedings of the 2016 International Symposium on Wireless Communication Systems (ISWCS), Poznań, Poland, 20–23 September 2016.

236. Hong, S.; Sagong, M.; Lim, C.; Cho, S.; Cheun, K.; Yang, K. Frequency and Quadrature Amplitude Modulation for Downlink Cellular OFDMA Networks. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1256–1267. [CrossRef]

237. Nekovee, M.; Yue, W.; Tesanovic, M.; Wu, S.; Qi, Y.; Al-Imari, M. Overview of 5G Modulation and Waveforms Candidates. *J. Commun. Inf. Netw.* **2016**, *1*, 44–60. [CrossRef]

238. Hochwald, B.M.; Sweldens, W. Differential Unitary Space-Time Modulation. *IEEE Trans. Commun.* **2000**, *48*, 2041–2052. [CrossRef]

239. Ostman, J.; Yang, W.; DUrisim, G.; Koch, T. Diversity Versus Multiplexing at Finite Blocklength. In Proceedings of the 11th International Symposium on Wireless Communications Systems (ISWCS), Barcelona, Spain, 26–29 August 2014.

240. Devaillie, B.; van Liempd, B.; Hershberg, B.; Craninckx, J.; Rikkinen, K.; van den Broek, D.J.; Klumperink, E.A.M.; Nauta, B. In-Band Full-Duplex Transceiver Technology for 5G Mobile Networks. In Proceedings of the ESSCIRC Conference, Graz, Austria, 4–8 December 2016.

241. Kim, C.; Kim, K.; Yun, Y.H.; Ho, Z.; Lee, B.; Seol, J.-Y. QAM-FBMC: A New Multi-carrier System for Post-OFDM Wireless Communications. In Proceedings of the IEEE Global Communication Conference GLOBECOM, San Diego, CA, USA, 6–10 December 2015.

242. Fettweis, G.; Krondorf, M.; Bittner, S. GFDM-generalized Frequency Division Multiplexing. In Proceedings of the IEEE 69th VTC, Barcelona, Spain, 26–29 April 2009.

243. Abdoli, G.; Jia, M.; Ma, J. Filtered OFDM: A New Waveform for Future Wireless Systems. In Proceedings of the IEEE 16th International Workshop SPAWC, Stockholm, Sweden, 28 June–1 July 2015.

244. Li, J.; Kearney, K.; Bala, E.; Yang, R. Filtered OFDM: A Resource Block Based Filtered OFDM Scheme and Performanse Comparison. In Proceedings of the 2013 20th International Conference on Telecommunications, Casablanca, Morocco, 6–8 May 2013.

245. Kumar, U.; Ibars, C.; Bhorkar, A.; Jung, H. Filtered OFDM: A Waveform for 5G: Guard Interval DFT-S-OFDM. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), San Diego, CA, USA, 6–10 December 2015.

246. Chung, C.D. Spectrally Precoded OFDM. *IEEE Trans. Commun.* **2006**, *54*, 2173–2185. [CrossRef]

247. Monk, A.; Hadani, R.; Tsatsanis, M.; Rakib, S. OTFS—Orthogonal Time Frequency Space. Available online: https://arxiv.org/ftp/arxiv/papers/1608/1608.02993.pdf (accessed on 10 July 2017).

248. Bandemer, B.; Gamal, A.E.; Kim, Y.H. Optimal Achievable Rates for Interference Networks with Random Codes. *IEEE Trans. Inf. Theory* **20125**, *61*, 6536–6549. [CrossRef]

249. Carleial, A.B. Multiple-access Channels with Different Generalized Feedback Signals. *IEEE Trans. Inf. Theory* **1982**, *28*, 841–850. [CrossRef]

250. Prasad, A.; Uusitalo, M.A.; Li, Z.; Lunder, P. Reflection Environment Maps for Enhanced Reliability in 5G Self-Organizing Networks. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–8 December 2016.

251. Sui, Y.; Papadogiannis, A.; Sternad, M. The Potential of Moving Relays-A Performance Analysis. In Proceedings of the 2012 IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, Japan, 6–9 May 2012.

252. Jiang, L.; Tian, H.; Xing, K.; Zhang, K.; Maharjan, S.; Cjessing, S.; Zhang, Y. Social-aware Energy Harvesting Device-to-Device Communications in 5G Networks. *IEEE Wirel. Commun.* **2016**, *23*, 20–27. [CrossRef]

253. Huang, X.; Yu, R.; Kang, J.; Gao, Y.; Maharjan, S.; Gjessing, S.; Zhang, Y. Software Defined Energy Harvesting Networking for 5G Green Communications. *IEEE Wirel. Commun.* **2017**, *24*, 38–45. [CrossRef]

254. Merwaday, A.; Guvenc, I. UAV Assisted Heterogeneous Networks for Public Safety Communications. In Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW), New Orleand, LA, USA, 9–12 March 2015.

255. Osseiran, A. The 5G Radio-Access Technologies. In *5G Mobile and Wireless Communications Technology*, 1st ed.; Osseiran, A., Monserrat, J.F., Marsch, P., Dohler, M., Nakamura, T., Eds.; Cambridge University Press: Cambridge, UK, 2016.

256. Sahin, A.; Yang, R.; Bala, E.; Beluri, M.C.; Olesen, R.L. Flexible DFT-S-OFD: Solutions and Challenges. *IEEE Commun. Mag.* **2016**, *54*, 106–112. [CrossRef]

257. Shi, S.; Yang, W.; Zhang, J.; Chang, Z. Review of Key Technologies of 5G Wireless Communication System. *MATEC Web Conf.* **2015**, *22*. [CrossRef]

258. Wang, C.-X.; Haider, F.; Gao, X.; You, X.-H.; Yang, Y.; Yuan, D.; Aggoune, H.; Haas, H.; Fletcher, S.; Hepsaydir, E. Cellular Architecture and Key Technologies for 5G Wireless Communication Networks. *IEEE Commun. Mag.* **2014**, *52*, 122–130. [CrossRef]

5G

VISION

DMC R&D Center, Samsung Electronics Co., Ltd.

Date : February 2015

Samsung envisions the fifth Generation (5G) mobile communication era to be the beginning of a full scale Internet of Things (IoT). Billions of connected devices autonomously interconnect with one another while ensuring personal privacy. The unprecedented latencies offered by 5G Networks will enable users to indulge in gigabit speed immersive services regardless of geographical and time dependent factors. This white paper introduces you to future services, key requirements, and enabling technologies that will herald in the 5G era that is expected to revolutionize the way we experience mobile services.

5G

Dawn of the 5G Era

Fuelled by the unprecedented growth in the number of connected devices and mobile data, and the ever-fast approaching limits of the 4G technologies to address this enormous data demand, industry efforts and investments to define, develop and deliver the systems and specifications for the fifth-Generation (5G) mobile system and services are well under way - signaling the dawn of the 5G Era.

As shown in Figure 1, the number of connected Internet of Things (IoT) is estimated to reach 50 Billion by 2020 [1], while the mobile data traffic is expected to grow to 24.3 Exabytes per month by 2019 [2].
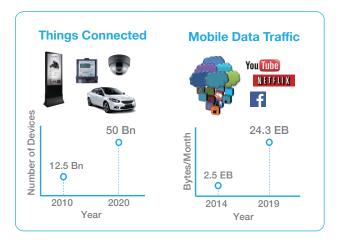
Figure 1   Growth in Mobile Traffic and Connected Devices

Add to it, the impact of higher cell capacity and end-user data rate requirements due to emerging new services such as Ultra-High-Definition (UHD) multimedia streaming and extremely low latency requirements for cloud computing and storage/retrieval, and it soon becomes evident that the current 4G systems, which are already stretched to near-breaking points (despite massive Wi-Fi offload), will be stretched too thin to deliver the quality of experience (QoE) necessary to support mobile experience that 5G is set to deliver.

5G Era can be expected to revolutionize the way we communicate by supporting immersive applications that demand exceptionally high-speed wireless connections, a fully-realized IoT, experience lower latency and promote both spectrum and energy efficiency. To realize these benefits, 5G systems will differ fundamentally from their predecessors fueling a series of groundbreaking innovations. Let's look at the services and the requirements that 5G is expected to address.

5G Service Vision

5G services have the potential to revolutionize the mobile experience. Here's how:

Internet of Things

5G will make the "Internet of Things" a reality. With 5G technology, a device will be able to maintain network connectivity regardless of time and location, and open the possibility to connect all the connected devices without human intervention. For this, the basic fabric of the 5G system design is expected to provide support for up to a million simultaneous connections per square kilometer, enabling a variety of machine-to-machine services including wireless metering, mobile payments, smart grid and critical infrastructure monitoring, connected home, smart transportation, and telemedicine. Intelligent devices will communicate with each other autonomously in the background and share information freely. This ubiquitous connectivity - a basic tenet of the 5G services, will truly enable IoT services which in turn is expected to profoundly change human lives by connecting virtually everything.

– Smart Home

Dishwashers will fix themselves using information from peers of the same model while home appliances at home and in the neighboring homes may cooperate to extinguish a fire. A smart refrigerator, recommending a recipe of cuisine to be cooked with ingredients that are already in your refrigerator, is yet another plausible scenario.

– Fitness & Healthcare

Connected Health and Fitness related wearable devices such as The Samsung Gear™ Fit will record your athletic performance while you exercise and recommend the type of exercise, its duration and frequency per day. These connected healthcare devices will also send vital signs such as brainwave, blood pressure and heartbeat to an expert system in the hospital in real-time to prevent medical emergencies before they occur. Such time critical applications put unprecedented requirements on latency.

– Smart Store

In large shopping malls, while many people walk around window shopping, vicinity to products is continuously tracked, usually by a server somewhere in the cloud. Customized alerts for low priced product can be sent to the user's device as the user is detected in the vicinity of that low-priced product, or information of other lower prices in the nearby stores can be sent to the device as the customer spends more time in the vicinity of a certain product class.

Such a system can be tailored to deliver a highly customized experience thereby greatly enhancing a user's shopping experience. To support such a scenario, massive connectivity and low latency technologies are inevitable.

— Smart Office

In smart office environments, office appliances are connected with one another and will share information. Nearby computers and input/output devices can recognize a user and change the settings using the user's preferences stored in the IoT cloud. Printers will print out the relevant documents when the user passes by the printer. Almost all the office appliances will connect wirelessly, while exchanging massive data through wireless medium without noticeable delay. Alerts on the upcoming meetings, materials and documents relevant to that meeting will instantaneously become available to the user's device, while documents and tasks that are modified will be automatically updated.

— Connected Car

Many of us use navigation services via in-car navigation systems or our smart phones to find the most efficient route to our destination. Vehicle diagnostic services are becoming attractive to obtain the information such as battery level, fuel level or engine status on our smartphones. The 'eCall' system that automatically calls emergency services in case of an emergency exemplifies such a service. By 2020 and beyond, more attractive services that wirelessly connect 'cars' around us with 'things' will emerge and make people in the vehicles very comfortable, and provide an enjoyable driving experience. For safer driving, sensor and camera data in a vehicle as well as supplementary information from the neighboring vehicles will be collected using mobile networks so that a potential emergency situation can be reliably informed to a driver in real-time and timely steps can be taken to avoid an actual emergency situation. This operation will eventually be applied to self-driving cars, which can be viewed as an important type of 'things'.

Immersive Multimedia Experience

In a 5G environment, users will experience life-like multimedia streams anytime and anywhere. Users will feel as if they are part of the scene when they watch videos on their smart devices. To provide such an immersive experience, many obstacles will need to be overcome. Agility to instantaneously respond to the user's thought and behavior will be necessary. An upcoming service that is expected to provide life-like experience in 5G system is UHD video streaming with its greatly enhanced resolution and clarity. Currently, UHD services over terrestrial broadcast are already being standardized in some countries. In addition, some smart phones in the market are now equipped with a camera module

Figure 2   Major Service Scenarios with 5G

that can record video with 4K UHD quality. It is expected that UHD services are likely to go mainstream by 2020 raising an acute need for performance enhancement of cellular systems to support such services.

Other examples of "immersive" services that will fundamentally revolutionize entertainment, health, education, and other industry sectors are Virtual Reality (VR) and Augmented Reality (AR).

VR provides a world where physical presence is simulated by computer graphics, and the user can actively interact with the simulated elements, as in immersive sports broadcasting for instance (See Figure 3). Other interesting VR service scenarios are interactive 360° movies, online games, remote education, and virtual orchestra. Samsung's recently launched VR headset called (called Gear VR), virtual reality video service platform (called Milk VR) and 'Project Beyond,' (a 360° 3D camera with 17 Full HD (FHD) camera modules, optimal for generating contents for Milk VR) indicate the humble beginnings of the truly immersive experience that is to come in the 5G Era.

Figure 3   Watching Sport Events with VR

In an AR service scenario, computer-aided real-time information based on user context is graphically augmented to the display, delivering added value for the user. In the future the service desk attendant no longer needs to memorize the tiresome details of the products. Instead, AR services will help to inform the price, popularity, and details of a given product. Figure 4 illustrates another service scenario - AR navigation on windshield, where navigation information and other helpful notifications (refuel reminder and nearby shop location) are displayed on the windshield of a car, so that the user can continue to focus on driving while getting subtle context-aware notifications about potential services at the same time.

Samsung is actively involved in the development of 5G technologies to support these immersive VR and AR services, which will entertain users and provide a truly life-like experience on the move.

Figure 4   Driving a Car with AR Navigation

Everything on the Cloud

5G will provide users with a desktop-like experience based on cloud computing. With everything stored and processed on the cloud and immediately accessed with low latency, only simple input and output interfaces on mobile devices are needed, making them lighter, thinner, fancier, and more eco-friendly.

As an example, when you go shopping, the smart device can notify you about the arrival of new dresses that you might like, or let you know how well a dress in the newly incoming inventory matches with your liking based on your purchase history. This notification can be triggered, for instance, as you step into a shop, or take a picture of the dress, while the necessary computation to come up with the dress recommendation through crowd sourcing is done on a cloud server that is potentially half a world away.

Intuitive Remote Access

Users will be able to control remote machines and appliances as if they are right before them, even from thousands of miles away. Thanks to the reliable connections and near-zero latency of 5G, users will be able to control heavy industrial machines remotely, or access hazardous site remotely. It will also help mankind in exploring areas on the earth that are as yet unexplored, such as the Polar Regions or parts of the ocean floor.

5G Requirements

In order to realize such a demanding and unprecedented service vision, Samsung proposes the 5G rainbow requirements consisting of 7 Key Performance Indices (KPIs) as shown in Figure 5.

5G systems will be required to deliver an order of magnitude cell capacities and per-user data rate compared to its predecessors. Specially, 5G systems are expected to support data rates of 10-50 Gbps for low-mobility users. As a baseline, 5G sys-
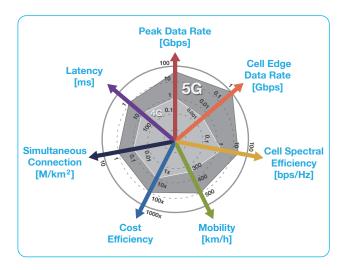
Figure 5   5G Rainbow Requirements

tems will provide gigabit-rate data services regardless of a user's location as shown in Figure 6 and Figure 7. To provide this uniform QoE, 5G network deployments are expected to be much denser compared to 4G networks, so making cost-effective deployment is a very important requisite.
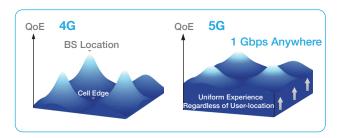
Figure 6   Edgeless RAN - 1 Gbps Anywhere
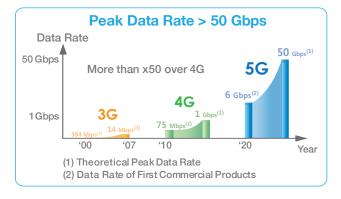
Figure 7   Data Rate Comparison of 5G with 3G and 4G

To fundamentally support the cloud storage/computing infrastructure of the future, 5G networks will deliver an end-to-end latency of less than 5 milliseconds and over-the-air latency of less than one millisecond (see Figure 8) - which is one-tenth compared to the latency of a 4G network. Critical infrastructure monitoring, for example, currently requires service levels achievable only on dedicated

wireline networks while 5G technologies offer the promise of making these service levels achievable over wireless networks. Likewise, low-latency networks will allow pre-crash sensing, enabling vehicles to sense imminent collisions and exchange relevant data that could salvage the situation and/or mitigate adverse impact of the collision. Other challenging low-latency services that could be enabled by 5G could include self-driving cars, public safety communications systems, augmented reality, and "tactile internet" [3].
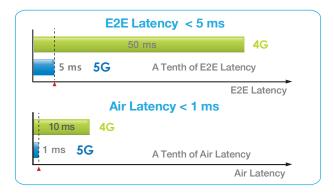
Figure 8   Ultra Low Latency of 5G

With spectral efficiency requirements set to 10 bps/Hz levels (in contrast to the 1-3 bps/Hz on 4G networks), 5G is also expected to deliver an efficient use of the spectrum by using MIMO, advanced coding and modulation schemes and new waveform design (more on this in the enabling technology section)

To address the ever-widening revenue gap that the operators and service providers are experiencing, 5G systems are targeted to be 50 times more efficient than 4G by delivering reduced cost and energy usage per bit. This sequentially requires low-cost network equipment, lower deployment costs, and enhanced power saving functionality on the network and user equipment sides.

5G technologies will be required to cope efficiently with all degrees of mobility by providing "mobility on demand" based on each device's and service's needs. On one hand, the mobility of user equipment should be guaranteed to be at least the same level as the current 4G system - that is the baseline. On the other hand, Samsung envisions that specialized 5G systems will support mobility at speeds ranging from 300 to 500 km/h.

To make the IoT Vision come true, the number of simultaneous connections in the 5G system is expected to be over 10^6 per unit square kilometer, which is much higher than that of the legacy system.

5G Key Enabling Technologies

Groundbreaking innovations will drive 5G technologies to meet the unprecedented speeds, near-wireline latencies, ubiquitous connectivity with uniform QoE, and the ability to connect massive amounts of devices with each other, all working in unison to provide the user with an immersive experience, even while the user is on the move. Future 5G systems will encompass fundamentally new designs to boost wireless capacity utilizing new frequency bands, advanced spectrum efficiency enhancement methods in the legacy band, and seamless integration of licensed and unlicensed bands.

Figure 9 shows an overview of the 5G key enabling technologies. The massively higher capacity needs of the 5G systems will be addressed by new mmWave systems, high-density small cells, advanced Multiple-Input and Multiple-output (MIMO) and new multiple access schemes like Filter-Bank Multi-Carrier (FBMC). Adaptive Coding and Modulation like Frequency and Quadrature Amplitude Modulation (FQAM) can significantly improve the cell edge performance and combined with higher density deployments with multi-BS cooperation will help to deliver on the promise of "Gbps anywhere" and Uniform QoE. Multi-Radio Access Technology (Multi-RAT) integration including carrier
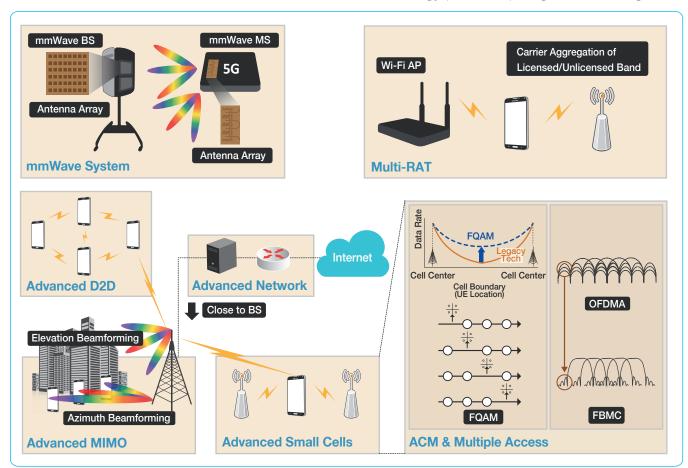
Figure 9 Overview of 5G Key Enabling Technologies

aggregation of licensed and unlicensed bands will inevitably help in increasing the available system bandwidth. On the network side, novel topologies including application servers placed closer to the network edge will contribute to significantly reducing the network latency. Advanced Device-to-Device (D2D) technology can help reduce the communications latency and support larger number of simultaneous connections in a network.

These 5G key enabling technologies are described in more detail in the following sections.

mmWave System

The mmWave bands provide 10 times more bandwidth than the 4G cellular-bands, as illustrated in Figure 10. Therefore, the mmWave bands can support the higher data rates required in future mobile broadband access networks.
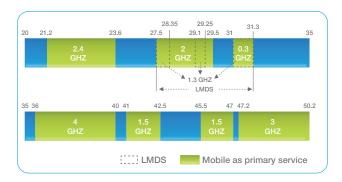
Figure 10   Potential Bands in 20-50 GHz (US)

Unlike the below 6GHz bands, in the mmWave bands we have to carefully consider the wireless conditions such as rainfalls, snowfalls, and fogs. However, the amount of loss is mild for the expected range of a 5G communication link. For example, atmospheric absorption loss due to H_2O and O_2 at 28 GHz would be around 0.02 dB at a distance of 200 meters. Intensive rainfall (110 mm/h) results in about 4 dB loss at 200 meters distance. Loss due to heavy snow (10 mm/h) or heavy fog (visibility distance under 50 meters) will be less than 0.1 dB, at all distance of 200 meters [4][5].

In Figure 11, we show mmWave channel measurements from extensive experiments in Daejeon, Korea [6][7]. These channel measurements were carried out at 28 GHz with a channel bandwidth of 250 MHz, transmission power of 29 dBm and the horn antenna gain of 24.5 dBi for both the transmitter and the receiver [6][7].

The path loss exponents (based on 8 meters distance free space path loss) is 3.53 in Non Line of Sight (NLoS) links [6]. The mild path loss exponent from measurement results strongly suggests that mmWave communication links can be supported

Panoramic View of Tx Position

Path Loss Model in Urban Environment
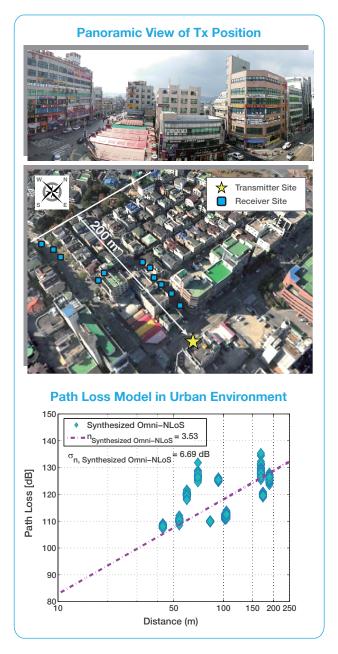
Figure 11 mmWave Channel Measurements in Urban Area, Daejeon (Korea)
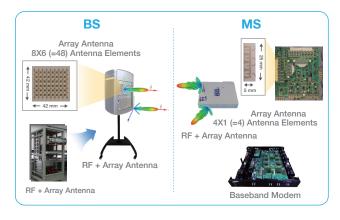
Figure 12   Adaptive Pencil Beamforming (Example)

over 200 meters of distance, even in dense urban outdoor NLoS environments. Similar results are also reported in ray-tracing-based mmWave propagation model [8], confirming the fact that mmWave

bands can be successfully utilized for future IMT systems. Moreover, based on the ray-tracing, mmWave 3D-channel model in urban scenario is proposed in [9]. Higher path-loss and the resulting fragile link is largely due to weak diffractions in the mmWave bands. We need to overcome these challenges in order to make outdoor mmWave communication a reality [10][11]. Fortunately, the small mmWave wavelength also means highly directional beamforms can be obtained using a large number of antenna elements in a smaller form factor. These adaptive directional beams with large antenna array gain are key in combating the large propagation loss in the mmWave [12][13][14][15], as illustrated in Figure 12.

We have developed a mmWave beamforming prototype at the DMC R&D Center, Samsung Electronics, Korea, in order to demonstrate the feasibility of using mmWave bands for cellular services. We showed that mmWave system can meet the two key requirements of cellular services: sufficiently large geographical coverage and support for mobility in NLoS environments.

With the prototype, we tested outdoor coverage to demonstrate the service availability in a typical outdoor environment for both LoS and NLoS sites. The tests were performed at sites surrounded by tall buildings where various channel propagation effects such as reflection, diffraction, and penetration were present, as shown in Figure 13. We observed that reliable communication links are formed even for NLoS sites that are more than 200 meters away from the base station.

Figure 13 Exemplary Outdoor Coverage Test Results

Advances in semiconductor technology have made commercial mmWave systems readily available. Recent technologies of Silicon-based Complementary Metal Oxide Semiconductor (CMOS) processes are capable of integrating mixers, Low Noise Amplifiers (LNAs), Power Amplifiers (PAs), and Inter-Frequency (IF) amplifiers, all in a single package. A good example is the 60 GHz commercialized products with the label of Wireless Gigabit Alliance (WiGig) [16], and it is well recognized that cost effective implementations of sub-100nm CMOS process made it possible to utilize the 60 GHz bands [17].

GaAs Monolithic Microwave Integrated Circuit (MMIC) technologies are maturing fast, and they are becoming a dominant choice for components in the RF chain including PAs, LNAs, switches for digital attenuators and phase shifters, Voltage Controlled Oscillators (VCOs), and passive components from a few GHz to 100 GHz. As illustrated in Figure 14, it is projected that the Power Added Efficiency (PAE) of a 5G Front End Modules (FEM) will soon match the PAE of today's commercial 4G system. Meanwhile, a good Effective Isotropic Radiated Power (EIRP) can be achieved with the help of high antenna gains from a large number of antenna elements [18][19].
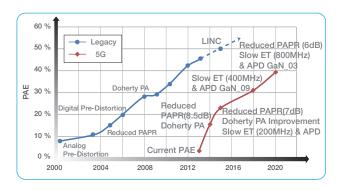
Figure 14 Projections of 5G FEM PAE

High performance antenna solution is an important piece of the 5G puzzle. The constantly varying LoS and NLoS propagation environment demands a novel antenna which exhibits high gain as well as wide spatial coverage capacity. The antenna array comprising of multiple antenna elements must fit within the small form factor of a 5G mobile handset.

Samsung has been developing innovative 5G phased array antennas that have near zero-footprint and reconfigurable antenna modes, as shown in Figure 15.
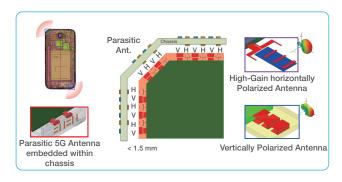
Figure 15 The Reconfigurable 5G Phased-array Antenna

Characterizing the biological implications on a user's body imposed by cellular handset devices is one of the most important aspects that need to be verified prior to its commercial release. The Signal Absorption Rate (SAR) regulated by worldwide gov-

ernmental bodies is used as a guideline to assess the effect of mmWave bands on a user's body. The SAR of an envisioned 5G (28 GHz) cellular handset is analyzed and illustrated in Figure 16.

When optimally configured, the 5G beamforming technology enables the peak radiation of the Mobile Station (MS) antenna to steer away from the user's head. As a result, the average SAR can be reduced by more than a factor of 10 compared to that of the present day 4G cellular handsets.

The maximum SAR is expected to be further minimized in the future, as 5G antenna technology continues to evolve.

■ 1.9 GHz (Omni)
: Penetration depth = 40 ~ 45 mm, Average = 0.29, Max =1 mW/g
Excluding the user's head | Including the user's head
Gain: 1 dBi
Penetration depth = 40 ~ 45 mm
Average/MAX = 0.29/1 (mW/g)

■ 28 GHz (Broadside Beamforming)
: Penetration depth = 3 mm, Average = 0.02, Max = 3.2 mW/g
Excluding the user's head | Including the user's head
Gain: 15 dBi
Penetration depth = 3 mm
Average/MAX = 0.02/3.2 (mW/g)
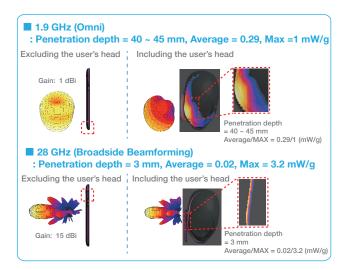
Figure 16   Biological Implications on the User's Body

Multi-RAT

To realize the envisioned 5G services, significant enhancement of per-user throughput and overall system capacity is required, compared with those of the 4G system. Such an enhancement can be achieved through advanced PHY/MAC/network technologies and efficient methods of cell deployment and spectrum management. In particular, utilizing a larger amount of system bandwidth guarantees an increase in the capacity by allocating more frequency resources to each user in the system. Therefore, utilizing the spectrum where huge bandwidth is available can be considered to be the most critical issue for the 5G system.

Currently, the 4G system specifies its operating frequency bands and some of them are assigned to cellular operators. These deployments can be beneficial since the existing 4G system will not interfere with other RATs.

However, obtaining the licensed spectrum requires not only considerable financial investment, but also a significantly long period of time spent on regulatory procedures. More importantly, a substantial portion of the licensed spectrum around 2 GHz is already being used. Therefore, finding a bandwidth

that is wide enough to support the 5G system is very challenging.

The recent trend in spectrum management is to aggregate both the licensed and unlicensed spectrums to extend available system bandwidth, as shown in Figure 17.
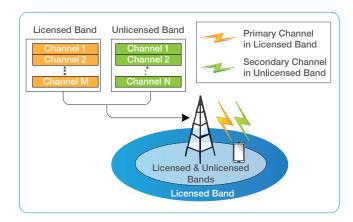
Figure 17   Integration of Licensed and Unlicensed Bands

The unlicensed spectrum has plenty of bandwidth. For instance, approximately 500 MHz and 7 GHz bandwidths are available in the 5 GHz and 60 GHz bands, respectively. In order for the 5G system to utilize the unlicensed spectrum, regulations imposed on each frequency band should be carefully reconsidered. Representative examples of the regulations are the Transmit Power Control (TPC), Dynamic Frequency Selection (DFS), and Listen Before Talk (LBT).

To efficiently utilize the unlicensed spectrum, we will develop the 5G system with the following characteristics.

First, we will design PHY/MAC/network algorithms suitable for the nature of the unlicensed spectrum. Since a wide range of frequency bands are included in the unlicensed spectrum, each frequency band has its own characteristics. Hence, band-specific solutions will be provided.

Second, efficient coexistence mechanisms that take into account other RATs (e.g., WiFi or WiGig) operating in the unlicensed spectrum will be suggested.

Finally, techniques for interworking and integrating the 5G system with other RATs will be developed. By taking advantage of multiple RATs, the 5G system will be able to take advantage of the unique characteristics of each RAT and improve the practicality of the system as a whole. For instance, the 4G system is used for exchanging the control messages to maintain the connection, to perform handover, and to provide real-time services such as VoLTE. The technology operating in mmWave unlicensed frequency band would support the gigabit data rate service. Multiple mmWave cells can be overlaid on top of the underlying 4G macro cells, as shown in Figure 18.
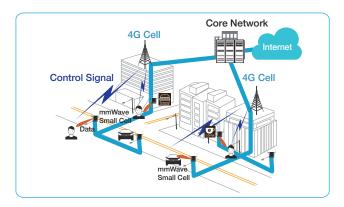
Figure 18   Overlaid Network of mmWave Small Cell Integrated with the Underlay 4G System

Advanced Network

In order to meet the 5G key requirements such as latency and the large number of simultaneous connections, and to support new business models and scenarios for operators, technologies above the radio access level should also be considered. Therefore, we need to develop technologies at the system architecture level from the network point of view. To accomplish the 5G key requirements and support the increased data rate of new 5G radio access, 5G network will have to evolve towards a distributed and flat architecture.

Current mobile network architectures designate a dedicated node in the core network (e.g., PGW - Packet Data Network Gateway in 3GPP) as a mobility anchor that allocates an IP address to the terminal, tracks terminal location in the IP topology, and ensures terminal's reachability by tunneling its traffic to wherever it goes. All terminal traffic is tunneled through the centralized node in the mobile core network. However, the undesirable consequences of this design include the following:

- Increase in end-to-end transmission latency due to elongated data paths.
- Additional load of backhauling and network processing in the core networks.
- Low network reliability due to introduction of a single point of failures.

In the 5G flat network architecture, as illustrated in Figure 19, user mobility is managed efficiently and in a dynamically scalable fashion by pushing the functionality to the edges of the network and even onto the mobile terminals [20].

The three important benefits of this approach include the followings:

First, such a distributed mobility management always provides the shortest data path between MS and the Internet without traversing the core network. This distributed mobility management leads to a

significant reduction of signaling and data transmission delay. Also, low end-to-end latency requirements of 'less than 5 ms' for new 5G services such as immersive UHD video streaming, cloud gaming, and virtual reality, cannot be archived solely by reducing radio access latency but would also require a flatter network architecture design. In the flat network architecture, services which require low latency transmission can be provided by Edge Servers and they can benefit from advanced network features which utilizes network information for optimal operations.

Second, it provides a highly scalable solution compared to the centralized architecture, in which a single core network gateway maintains the whole traffic from MSs or to MSs.

Third, it easily avoids the risk of having a single point of failure. In flat network architecture, the breakdown of one network gateway would not significantly interfere with the operations of the other gateways.

Flexibility is considered as another key requirement of 5G network architecture. Software-Defined Networking (SDN) and Network Function Virtualization (NFV) provide promising examples of programmable design technologies for realizing a flexible 5G network architecture.
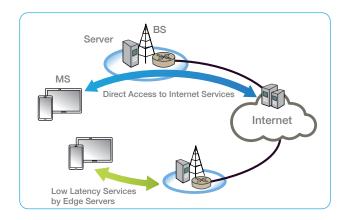
Figure 19   5G Flat Network Architecture

Advanced MIMO

One promising technology for meeting the future demands is massive MIMO transmission/reception [21]. When used with multi-user precoding schemes such as Maximum Ratio Transmission (MRT) precoding, also known as channel conjugate precoding, massive MIMO systems experience small inter-user and inter-cell interferences, and consequently achieve significantly higher throughput than the state-of-the-art MIMO systems.

One of the main challenges to build massive MIMO systems in practice is the limitation in the number of antennas that can be equipped at a BS, caused by the BS form factors and operating carrier frequencies.

For example, to horizontally install a large number of antenna elements (e.g., > 8) at the top of a BS tower operating with the lowest 4G system frequency bands of 700 MHz, eight antenna elements with 0.5λ spacing require up to 1.7 m width, where λ is the carrier wavelength. For the typical 4G system frequency bands of 2.5 GHz, fitting 32 antenna elements with 0.5λ spacing require up to 1.9 m width, which is still not feasible in many BSs that have only limited room on the tower. This practical limitation has motivated Full-Dimension MIMO (FD-MIMO) cellular communication systems, which place a large number of active antenna elements in a two dimensional grid at the BSs.

A typical FD-MIMO deployment scenario is illustrated in Figure 20, for a macro BS with 3 sectors equipped with 2D Active Antenna Array (AAA) panels.
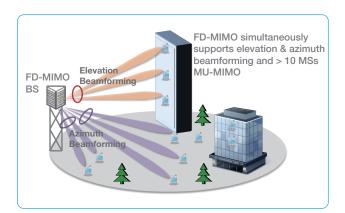
Figure 20   Example of FD-MIMO Deployment

FD-MIMO system can support high-order Multi-User MIMO (MU-MIMO), while fully exploiting the elevation as well as the azimuth dimension, thereby generating improved system throughput.

In full buffer system-level evaluations, it has been found that 64-antenna-port FD-MIMO system achieves 243% average-cell and 244% cell-edge performance gain, compared to those of the 8-antenna-port legacy MIMO system. In order to achieve the promising gain of FD-MIMO in practice, we need accurate beam steering and tracking in three Dimensions (3D).

To steer and track the MU beams toward serving MSs, FD-MIMO BS should be equipped with multiple transceivers (TRX) feeding 2D array elements, in which the number of TRX doubles or even quadruples compared to that of the conventional BS.

Having a large number of TRX poses new challenges, such as antenna calibration and complexity issues associated with Channel State Information (CSI) acquisition and precoding.

On the other hand, high-order MU-MIMO introduces another set of new challenges, such as scheduling complexity and link adaptation. Furthermore, in Frequency Division Duplex (FDD) systems, other new challenges emerge such as pilot overhead, CSI estimation complexity, CSI quantization and feedback overheads.

ACM & Multiple Access

– FBMC

As cellular IoT has been one of key driving forces to 5G, spectrally efficient support for heterogeneous services that have quite different requirements is becoming ever so important. Accordingly, several enabling methods such as multi-RAT coexistence and flexible spectrum sharing have been actively investigated.

Recently, FBMC has drawn much attention as an enabling technology for enhancing the fundamental spectral efficiency, though its theory has a long history similar to that of Orthogonal Frequency Division Multiplexing (OFDM).

Because of the well-localized time/frequency traits adopted from a pulse shaping filter per subcarrier, the FBMC system can reduce the overhead of guard band required to fit in the given spectrum bandwidth, while meeting the spectrum mask requirement.

Furthermore, the effectively increased symbol duration is suitable for handling the multi-path fading channels even without Cyclic Prefix (CP) overhead. Consequently, the FBMC system can reduce the inherent overheads such as CP and guard-bands in CP-OFDM. FBMC is also attractive in specific asynchronous scenarios, including Coordinated Multi-Point Transmission and Reception (CoMP) and Dynamic Spectrum Access (DSA) in a fragmented spectrum.

However, to maintain the transmission symbol rate, the conventional FBMC system generally doubles the lattice density either in time or in frequency compared with OFDM while adopting Offset Quadrature Amplitude Modulation (OQAM). In OQAM, in-phase and quadrature-phase modulation symbols are mapped separately with half symbol duration offset. Thus, so-called OQAM-FBMC or Staggered Multi-Tone (SMT) causes intrinsic interference that makes it difficult to apply conventional pilot designs and corresponding channel estimation algorithms as well as MIMO schemes as in CP-OFDM systems [22].

With a base-filter that takes the spectrum confinement and the orthogonality among adjacent subcarriers into consideration, the QAM-FBMC system performs comparable to the CP-OFDM system even without the CP overhead, while the guard-

band overhead reduction is also available from the well-confined spectrum. Sophisticated receiver algorithms including channel estimation and equalization can further mitigate the multi-path fading channel impact without the CP.
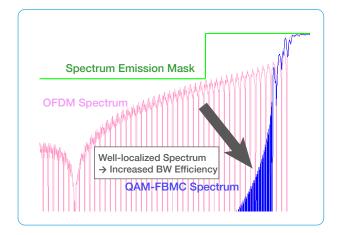
Figure 21   One QAM-FBMC Symbol Generation

– **FQAM**

One of key requirements for 5G is enhancement of cell-edge performance, which means that every user should be supported with Gigabit experience anywhere. Conventional approaches to enhance the cell-edge performance mainly focus on managing interference (e.g., interference cancellation, interference avoidance), by dealing with interference as a Gaussian. However, it is proved that the worst-case additive noise in wireless networks with respect to the channel capacity has a Gaussian distribution. From this observation, one can expect that the channel capacity can be increased by a non-Gaussian interference design which makes Inter-Cell Interference (ICI) non-Gaussian. The distribution of ICI depends on the modulation schemes of the interfering BSs. Therefore, an active interference design for improved cell-edge performance can be achieved by applying a new type of modulation.

FQAM, a combination of Frequency Shift Keying (FSK), and Quadrature Amplitude Modulation (QAM) can be used as an active interference design scheme. Figure 22 shows the signal constellation of 16-ary FQAM that is a combination of 4-ary FSK and 4-ary QAM.
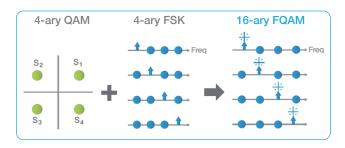
Figure 22   Example of 16-ary FQAM

With FQAM, the statistical distribution of ICI is likely to be non-Gaussian, especially for cell-edge users. As a result, the transmission rates for the cell-edge users can be significantly improved.

The statistics of ICI and the performance enhancement possibility have been proven by practical implementation of a system which uses FQAM. FQAM system environment for cellular downlink OFDMA networks is shown in Figure 23 and 24. Our experimental results show that the transmission rates for interference-limited users in FQAM-based OFDMA networks are around 300% higher than those in QAM-based OFDMA networks [23][24].

Figure 23   3 Cell Structure with MS and BSs
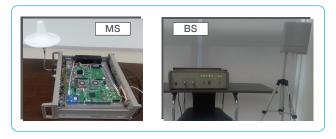
Figure 24   Implemented MS and BS

Advanced D2D

Advanced D2D communication is an attractive technology that enhances spectral efficiency and reduces end-to-end latency for 5G. Not entirely depending on the cellular network, D2D devices can communicate directly with one another when they are in close proximity. Hence, D2D communication will be used for offloading data from network so that the cost of processing those data and related signaling is minimized. Moreover, enhanced version of D2D communication was suggested lately for being used as special purpose such as Mission Critical Push-To-Talk (MCPTT) communication and Vehicle-to-Anything (V2X) communication.

In advanced D2D communication, a single radio resource can be reused among multiple groups which want to communicate with each other if the interference incurred between groups is tolerable. Hence, we can increase the spectral efficiency and the number of the simultaneous connection by utilizing D2D communication in 5G. Moreover, since the data is directly transmitted and not going through the core network, the end-to-end latency can be

considerably reduced. Therefore, advanced D2D communication is quite well aligned with IoT services as shown in Figure 25. The cars can communicate with each other to exchange the information for safety alarm and infotainment without cellular base station. The home appliances communicate with each other for home automation service. Many objects in proximity region can be connected to each other so that IoT services can be accomplished.
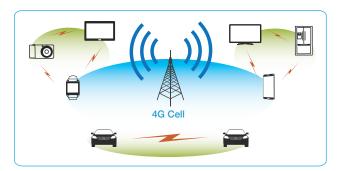
Figure 25   Advanced D2D Communications

Advanced Small Cell

To achieve significant throughput enhancement in a practical manner, it is necessary to deploy a large number of cells in a given area and to manage them intelligently. The 5G system is expected to utilize higher frequencies to take advantage of the vast bandwidth in the mmWave bands. Hence, the considerably high propagation loss of mmWave makes it suitable for dense small cell deployment, which leads to higher spatial reuse.

Moreover, Figure 26 shows the concept of a user-centric virtual cell. Conventional static network topologies with a central controller have an "edge", the reach of the central controller. However, a user-centric virtual cell that consists of a group of cooperating BSs is continuously reformed so that any user will always find himself/herself at the "center" of the cell.

Distributed and self-configuring network technologies will make it easy to deploy many small BSs in urban and suburban areas. In-band wireless backhaul can be used between BSs for cooperative communication, reducing the cost and complexity of backhaul network deployment.
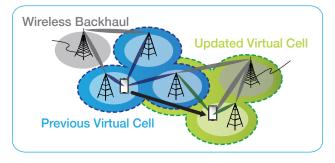
Figure 26   User Centric Virtual Cellular Network

5G Deployment Scenario

In order to provide the ubiquitous high QoE gigabit accessibility, we envision an overlaid deployment of 5G in conjunction with the existing 4G macro cells. The 5G small cells can be coupled with the overlaid 4G macro cells, and the 4G macro cells will control the operations of the associated 5G small cells, as illustrated in Figure 27. In the figure, 5G BS primarily provides multi-gigabit per second throughput with high QoE to mobile users over the legacy spectrum or higher spectrum like the 5 GHz unlicensed band and bands above 6 GHz. Meanwhile, the 4G BS can serve as a control channel to 5G MSs for supporting seamless connection anywhere over the legacy 4G spectrum.

5G system will need enhanced RAN technologies that not only utilize the legacy frequency bands assigned to IMT and IMT-Advanced systems, but also use new frequency bands. At the same time, 5G system will also need to support seamless interworking with other RATs operating in unlicensed bands.
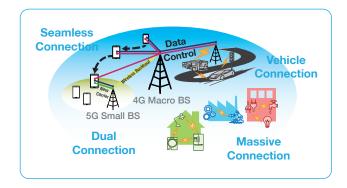
Figure 27   5G Deployment Scenario

Standardizations and Regulations

As shown in Figure 28, 5G is planned to be commercialized in year 2020. ITU-R WP 5D is preparing timelines in the vision document for standardization, spectrum allocation, and commercialization, in order to be on time. From the perspective of commercialization, standards for 5G should be ready by the year 2017 to allow 2 or 3 years for the development of 5G products. Considering average periods of previous standardization, 5G standards need to get started in 2015 to make 5G standards available by the end of 2017.

New frequency bands may be required to achieve the target performance of 5G. World Radio communication Conference (WRC)-15 agenda item 1.1 made in WRC-12 indicates consideration for additional spectrum allocations to mobile services on a primary basis and for identification of additional frequency bands for IMT purposes.
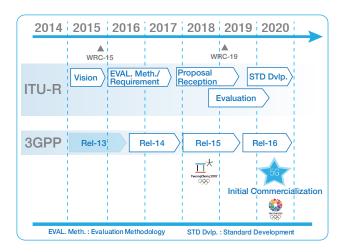
Figure 28   5G Timelines

In light of this agenda item, various frequency ranges had been proposed and discussed in ITU-R WP 5D in July 2013, and Joint Task Group (JTG) 4-5-6-7 had performed compatibility studies among various technologies in each candidate frequency band until July 2014.

Although bands above 6 GHz were not included in JTG 4-5-6-7's candidate frequency bands, WP 5D has been developing a new report on IMT systems in bands above 6 GHz, including channel characteristics and the potential usage of the bands for IMT systems.

WP 5D may also provide JTG 4-5-6-7 with the necessary information regarding bands above 6 GHz in order to be referenced as appropriate in the draft text of Conference Preparatory Meeting (CPM). The possibility and further necessity to study the bands are captured in the report from JTG 4-5-6-7 to CPM as follows:

"The demand of high bit rate, especially in densely populated area could be accommodated in higher frequency bands (e.g. above 6 GHz) than those currently being considered in studies, however the technical information required for compatibility studies has yet to be developed and these studies and proposals are being explored for future work, beyond WRC-15."

This report implies that new frequency bands can be added to the IMT frequency bands toward WRC-19, and it is likely that bands above 6 GHz can be one of candidate frequency bands.

In addition, Licensed Assisted Access (LAA) study item has been accepted in 3GPP LTE in the second half of 2014, acknowledging the use of unlicensed bands for cellular system.

Recently, as FCC relieves 5GHz band regulations for various unlicensed band usage, related coexistence technology between LTE and WLAN will be standardized in both standards.

Further, as shown in Figure 29, Samsung is actively engaging in the most of global 5G research initiatives, such as European 5G PPP projects of Horizon 2020, 5G Innovation Centre (5GIC) in UK, NYU Wireless Center in US, Giga KOREA project and Chinese 836 project.

Samsung is leading various collaborations with industries and academics over the world. In particular, Samsung has played an important role as the full member of 5G PPP Infrastructure Association, the executive board member of 5G Forum in Korea and the chair of vision sub-working group for Future IMT (5G) in ITU-R WP5D.

In order to have a consistent perspective on 5G with those of other academic institutes, we are vigorously developing 5G core technologies with several outstanding universities around the world.

Figure 29   Global 5G R&D Activities

Conclusion

5G will usher in a revolutionary generation of mobile communication that provides ubiquitous multi-Gbps data rate regardless of the user's location.

Significantly increased system capacity and real-time responsiveness of the 5G system will introduce life-changing services providing the users with a truly immersive and rich experience. The profoundly life-alternating world of billions of interconnected devices, the IoT, will require something as revolutionary as 5G to be fully realized in its fullest extent. And finally 5G promises to reverse the widening revenue gap and make it worthwhile for operators and service providers to invest again in innovative new services, and continue to propel increased productivity and efficiency.

Reference

1. UMTS, "Mobile Traffic Forecasts 2010-2020 Report," UMTS Forum, Jan. 2011.

2. Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update: 2013-2018," Cisco, Feb. 2014.

3. Calum Dewar et al., "Understanding 5G: Perspectives on future technological advancements in mobile," GSMA Intelligence Report, Dec. 2014.

4. ITU-R P.840-5, "Attenuation Due to Clouds and Fog," ITU-R, Feb. 2012.

5. DARPA R-1335-ARPA, "Atmospheric Effects on Terrestrial Millimeter-Wave Communications," DARPA Report, Mar. 1974.

6. Sooyoung Hur et al., "Wideband Spatial Channel Model in an Urban Cellular Environments At 28 GHz," in Proc. EuCAP'15, April. 2015.

7. Sooyoung Hur et al., "28 GHz Channel Modeling Using 3D Ray-Tracing in Urban Environments," in Proc. EuCAP'15, April. 2015.

8. Youngbin Chang et al., "A Novel Dual-Slope mm-Wave Channel Model Based on 3D Ray-Tracing in Urban Environments," in Proc. IEEE PIMRC'14, Sep. 2014.

9. Sooyoung Hur et al., "Synchronous Channel Sounder Using Horn Antenna and Indoor Measurements on 28 GHz," in Proc. IEEE BlackSeaCom'14, pp. 83-87, May 2014.

10. Zhouye Pi et al., "An Introduction to Millimeter-Wave Mobile Broadband Systems," IEEE Communications Magazine, vol.49, no.6, pp.101-107, Jun. 2011.

11. Cheol Jeong et al., "Random Access in Millimeter-Wave Beamforming Cellular Networks: Issues and Approaches," IEEE Communications Magazine, vol. 53, no. 1, pp.180-185, Jan. 2015.

12. Wonil Roh et al., "Millimeter-wave Beamforming as an Enabling Technology for 5G Cellular Communications: Theoretical Feasibility and Prototype Results," IEEE Communications Magazine, vol. 52, no. 2, pp. 106-113, Feb. 2014.

13. Samsung, "Technologies for Rel-12 and Onwards," RWS-120021, 3GPP TSG RAN Workshop on Rel-12 and Onwards, Jun. 2012.

14. Taeyoung Kim et al., "Tens of Gbps Support with mmWave Beamforming Systems for Next Generation Communications," in Proc. IEEE GLOBECOM'13 Workshop, pp.3685-3690, Dec. 2013.

15. Chanhong Kim et al., "Multi-Beam Transmission Diversity with Hybrid Beamforming for MIMO-OFDM Systems," in Proc. IEEE GLOBECOM'13 Workshop, pp. 61-65, Dec. 2013.

16. WiGig Alliance, "Defining the Future of Multi- Gigabit Wireless Communications," WiGig White Paper, Jul. 2010.

17. R12-WP5D-C-0258, "Technical Feasibility of IMT in the Bands above 6 GHz," ITU, Jan. 2013.

18. Theodore S. Rappaport et al., "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," IEEE Access, vol.1, pp.335-349, May 2013.

19. Theodore S. Rappaport et al., "28GHz Propagation Measurements for Outdoor Cellular Communications Using Steerable Beam Antennas in New York City," in Proc. IEEE ICC'13, pp.5143-5147, Jun. 2013.

20. Alper Yegin et al., "Terminal-centric Distribution and Orchestration of IP Mobility for 5G Networks," IEEE Communications Magazine, vol. 52, no. 11, pp.86-92, Nov. 2014.

21. Young-Han Nam et al., "Full Dimension MIMO (FD-MIMO) for Next Generation Cellular Technology," IEEE Communications Magazine, vol. 51, no. 6, pp.172-179, Jun. 2013.

22. Maurice Bellanger et al., "FBMC Physical Layer: a Primer," PHYDYAS Tech. Rep., Jun. 2010.

23. Sungnam Hong et al., "A Modulation Technique for Active Interference Design under Downlink Cellular OFDMA Networks," in Proc. IEEE WCNC'14, pp.683-688, Apr. 2014.

24. Sungnam Hong et al., "Frequency and Quadrature-Amplitude Modulation for Downlink Cellular OFDMA Networks," IEEE Journal on Selected Areas in Communication, vol. 32, no. 6, pp.1256-1267, Jun. 2014.

Samsung Demonstrates World's First 5G Data Transmission at Highway Speeds

Record-breaking 1.2Gbps data transmission at over 100km/h, and 7.5Gbps in stationary conditions using 28GHz spectrum

SEOUL, Korea – **October 15, 2014** – Samsung Electronics Co., Ltd today announced the record-breaking demonstration of super-fast data transmission targeted for fifth generation (5G) mobile networks using 28 GHz spectrum. Samsung researchers confirmed the world's first data rate of 1.2 Gbps, or 150 MB per second on a vehicle cruising at over 100 km/h. This marks a significant step towards the utilization of millimeter wave frequency bands for 5G mobile networks

Figure A   Achieved Data Rate in Mobile Condition (1.2Gbps)

The achievement was bolstered by another record-breaking demonstration in which Samsung achieved data transmission speed up to 7.5 Gbps, or 940 MB per second when the vehicle came to a complete stop. The peak data rate is more than thirty times faster compared to the state-of-the-art 4G LTE technology.

Back in May, 2013, Samsung revealed the world's first 28 GHz based 5G data transmission speed of 1Gbps at pedestrian speeds. Since then researches at Samsung were able to increase the maximum data rate by more than 7-fold and support mobility up to highway speeds. The continued ground breaking success underlines the company's leadership in next-generation mobile communications.

"The expectations for 5G communications will continue to grow based on the rising demands for smart devices, cloud services, smart home technology, and Internet of Things" , said ChangYeong Kim, Head of DMC R&D Center of Samsung Electronics. "We are committed to developing innovative 5G technologies and will continue to utilize our exceptional R&D capabilities as well as our diverse partnerships with leading companies and research centers around the world".

While 5G standard has yet to be ratified, 5G networks are expected to feature data rates and capacity that are hundreds times larger compared to 4G LTE.

Figure B   Achieved Data Rate in Stationary Condition (7.5Gbps)

In order to achieve such high data rates, 5G networks will inevitably exploit frequencies much higher and less congested than those of which are currently used for cellular networks (typically under 3 GHz). However, difficulties such as large propagation loss at these frequency bands have prohibited the industry from utilizing these bands for cellular applications. To address these challenges, Samsung has applied 'Hybrid Adaptive Array Technology' at 28 GHz frequency bands.

In addition, Samsung has revealed its '5G Rainbow' – seven technical requirements which are pillars to ensure a truly differentiating 5G user experience. The list comprises of maximum data rate, spectral efficiency, speed of mobility, data transmission rate at the cell boundary, the number of simultaneous connections, communication delays, and cost. To meet these requirements, Samsung has been developing key technologies including transmission technologies for high frequency bands, multiple access schemes and low latency networks among others.

The company is committed to developing innovative technologies that will enable the 5G era and plans to demonstrate its capabilities at the upcoming 2018 PyeongChang and 2020 Tokyo Olympics.

Australian Government
Department of Communications and the Arts

5G—Enabling the future economy

October 2017

Disclaimer

The material in this paper is of a general nature and should not be regarded as legal advice or relied on for assistance in any particular circumstance or emergency situation. In any important matter, you should seek appropriate independent professional advice in relation to your own circumstances.

The Commonwealth accepts no responsibility or liability for any damage, loss or expense incurred as a result of the reliance on information contained in this paper.

This paper has been prepared for consultation purposes only and does not indicate the Commonwealth's commitment to a particular course of action. Additionally, any third party views or recommendations included in this paper do not reflect the views of the Commonwealth, or indicate its commitment to a particular course of action.

Copyright

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Contents

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Australia's 5G vision

5G is the next step in the evolution of mobile wireless communications technology, promising improved connectivity, greater network speeds and bandwidth, and very low latency. It is the fifth generation in mobile technology which, at each step, has seen significant developments in communications networks:

- 1G—The first generation of mobile phone networks were deployed in the early 1980s, providing a basic voice service using analogue transmission.
- 2G—In 1991, second generation networks were deployed, making the switch to digital standards with improved voice messaging and the introduction of the short message service (SMS).
- 3G—The third generation launched in 2001 and introduced data services in addition to voice and SMS.
- 4G—In 2009, the fourth generation protocol, Long Term Evolution (LTE), was introduced, supporting improved mobile broadband which saw increased capacity and speed for data.

New capabilities of mobile communications networks enabled by 5G technology will allow for a variety of 'use cases':

- higher quality and more video services provided to multiple users with full mobility, even at high speed
- massive scale automation delivered through widespread sensor networks and multiple connected devices
- delivery of critical communications assured by low latency and ultra-reliable networks, and
- improved productivity assisted by high quality, real time data analytics.

Unlike existing mobile communications networks, 5G networks have the potential to allow tailoring of requirements for each of these different use cases within the same network.

The Government considers that 5G is more than an incremental change for mobile communications. Instead, it provides the underlying architecture that will enable the next wave of productivity and innovation across different sectors of the Australian economy. Efficient rollout of 5G and uptake of the services it supports has the potential to produce far-reaching economic and social benefits and support growth of Australia's digital economy. This will be supported by the rollout of the National Broadband Network (NBN) allowing greater capabilities for the seamless delivery of services across high speed mobile, fixed line and fixed wireless networks.

The Government wants to create an environment that allows Australia's telecommunications industry to be at the forefront of seizing the benefits of 5G across the economy. The communications sector will lead the rollout of 5G networks in Australia. However, the Government can create the policy and regulatory environment to support a more efficient rollout, given its potential benefits to the economy.

The Government's direction will be to support the timely rollout of 5G in Australia to enable the next wave of broad-based industry productivity, and support the growth of Australia's digital economy.

This includes immediate actions by Government that enable the communications market to introduce new 5G technologies in line with international developments. These include:

- making spectrum available in a timely manner
- actively engaging in the international standardisation process
- streamlining arrangements to allow mobile carriers to deploy infrastructure more quickly, and
- reviewing existing telecommunications regulatory arrangements to ensure they are fit-for-purpose.

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 1 of 12

The Government recognises that as 5G continues to develop, other issues relating to the technology will likely emerge which may require future Government action. In particular, while there are opportunities for 5G to create economy-wide transformation, this will require a broader examination of sectoral regulatory frameworks.

To that end, the Government will work collaboratively with industry to foster an ongoing dialogue on 5G beyond the launch of this paper to identify and remove sectoral barriers to its successful and timely rollout. Through this dialogue, the Government will also look at opportunities to build on other Government activities, such as the national Digital Economy Strategy which will more broadly focus on building the productivity of sectors across the economy.

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 2 of 12

What is 5G?

The International Telecommunication Union (ITU) is the United Nations specialised agency for information and communications technologies. This body decides global spectrum allocation frameworks and harmonises international spectrum to ensure networks and connected devices can communicate seamlessly. The ITU will undertake the formal, international process to identify bands for 5G by 2020. It has developed draft technical specifications for 5G which include:[1]

- high data rates (1 Gbps for hotspots, 100 Mbps download and 50 Mbps upload for wide-area coverage)
- massive connectivity (1 million connections per square kilometre)
- ultra-low latency (1 millisecond)
- high reliability (99.999% for mission critical 'ultra-reliable' communications), and
- mobility at high speeds (up to 500 km/h i.e. high speed trains).

In working towards these specifications, 5G represents a significant leap from the capabilities of previous generations and introduces a range of new technological possibilities. The success of 5G in delivering new technologies and services will be supported by existing communications infrastructure, including the NBN. This convergence of high-speed fixed-line and mobile services will collectively produce a consistent and ubiquitous user experience.

The 5G economics case

Unlike early generations of mobile networks, 5G will represent a significant shift in the telecommunications industry's focus away from voice and more towards mobile broadband and increased industrial applications. These new use cases are expected to create benefits across a range of sectors—including transportation, health, manufacturing and agriculture—and have varying networking requirements. These use cases, as identified by the industry, can be divided into the following categories:

- enhanced Mobile Broadband
- massive Machine Type Communications, and
- critical communications.

Enhanced Mobile Broadband (eMBB)

eMBB will deliver improved capacity to a greater number of devices. This will enable higher rates and volumes of data transmission per device and improve coverage to a broader range of locations. eMBB will likely be the focus of early 5G deployments as it can immediately support the growing communications requirements for the digital economy.

An improved mobile experience for consumers

5G networks will give consumers a better mobile experience in more locations. Increased network capacity will support more users, even in crowded areas, such as large public events, and at peak times. Faster network speeds will also enable consumers to view rich content in more places, supporting the streaming of live events and high resolution media.

---

[1] ITU, 23 February 2017, 'Press Release: ITU agrees on key 5G performance requirements for IMT-2020'.

5G—Enabling the future economy   www.communications.gov.au
www.arts.gov.au   Page 3 of 12
www.classification.gov.au

Massive machine type communications (mMTC)

As 5G networks mature, they will support the widespread and dense deployment of sensors and other network-connected devices by significantly reducing their power requirements and providing flexible coverage across different spectrum bands. This proliferation of the Internet of Things (IoT) across industries is expected to produce significant productivity benefits and support integration between sectors.

Supporting productivity and innovation

The term Industry 4.0 describes the next step in the advancement of the manufacturing sector (the 'fourth industrial revolution'). Industry 4.0 introduces autonomous systems supported by a combination of technologies such as IoT, artificial intelligence, continued technological improvements and digitalisation in manufacturing.

Australia stands to benefit from Industry 4.0, given our world-class manufacturing sector, which includes several high-value industries such as medical technology and aerospace. Australian manufacturers can improve their productivity and international competitiveness through Industry 4.0 processes by supporting their participation in global value chains. This is of particular benefit to SME manufacturers, opening them up to new markets and opportunities. 5G can support Industry 4.0, by providing communications infrastructure that is more accessible and flexible to suit specific industry needs.

5G can enable innovation in other sectors such as agriculture. A challenge for Australia's agricultural sector is identifying how to improve productivity while balancing environmental and commercial constraints. Precision agriculture, which focuses on improving yields and minimising economic risks, seeks to provide more control in the management of agricultural production. While precision agriculture requires a range of enablers—including data analysis, sensor networks and geographical information systems—5G can provide the supporting infrastructure for these technologies.

Critical communications

Low latency and 'ultra-reliable' communications networks will support the delivery of critical communications, i.e. to support public safety use and playing role in the technology ecosystem supporting autonomous vehicles. In addition to automation, critical communications will also help to support technological advancement in areas including robotics and artificial intelligence.

The social benefits from autonomous vehicles

5G networks are expected to play a role in the technology ecosystem supporting the development of autonomous vehicles, which will enable a number of social benefits for transportation. Traffic congestion, which is estimated to cost Australia $53 billion by 2031,[2] could be proactively reduced by smart city traffic management systems that are informed by machine-to-machine communications with autonomous vehicles.
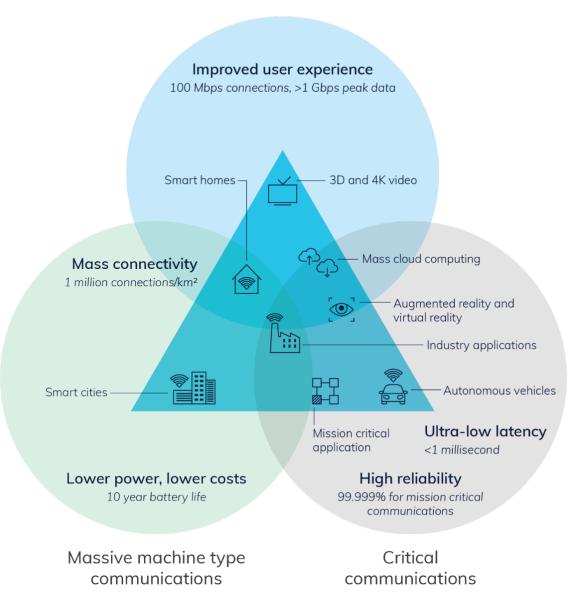
Improved road safety is also expected to be a key outcome of autonomous vehicles, as the majority of car accidents involve human error. In the 12 months to July 2017, there were 1,235 deaths on Australian roads with road trauma costing the Australian community an estimated $27 billion annually.[3] Autonomous vehicles can have a valuable role not just in terms of financial savings, but in saving human lives.

---

[2] Infrastructure Australia, 2015, 'Australian Infrastructure Audit'.
[3] Bureau of Infrastructure, Transport and Regional Economics, August 2017, 'Road deaths—monthly bulletins'.

www.communications.gov.au
5G—Enabling the future economy   www.arts.gov.au   Page 4 of 12
www.classification.gov.au

Figure 1—Outline of the relationship between the technical requirements of 5G and expected services and applications it will deliver

Source: Based on figure 2 from the ITU paper, 'IMT Vision—Framework and overall objectives of the future development of IMT for 2020 and beyond.'

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 5 of 12

What is different about 5G?
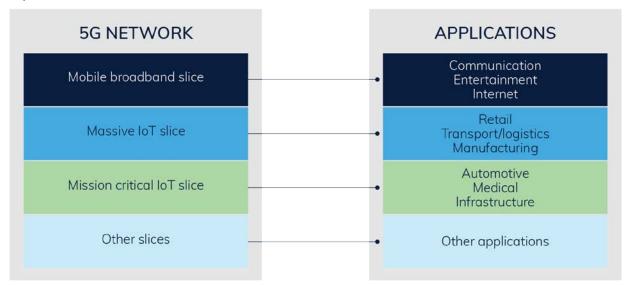
5G is expected to enable productivity outcomes across key verticals of the economy as a result of a range of characteristics, such as **network slicing** and **mesh networks**. However, it will need to overcome deployment challenges such as **spectrum acquisition** and **coverage**. This is expected to lead to new approaches for deployment to enable widespread 5G coverage in Australia. The industry will also need to develop a robust **business case** for 5G rollout, which will likely be demand driven, to support the significant capital expenditure required.

Network slicing

Network slicing allows operators to split their network into separate sub-networks (also referred to as slices), enabling them to dedicate network resources to different users and applications. Sub-networks can 'slice' the infrastructure resources from the physical network to create virtual independent networks. This is a significant development from previous mobile network generations, as it enables an operator to deliver many different capabilities by creating slices that can be tailored for the intended usage. For example, an operator could create a network slice for IoT devices, or alternatively, a network slice for higher security for a higher quality of service for government or public safety uses.

Network slicing also provides a model for infrastructure sharing between network operators. In this case, a single transmitter's network could be shared with more than one operator. Given the capital expenditure that will be associated with 5G rollout, network slicing can provide a cost effective, short-term solution for operators.

Figure 2—5G network slicing enables operators to create separate virtual networks to meet application requirements

Source: Based on figure 2 from the ITU article, '[Why end-to-end network slicing will be important for 5G](#).'

Mesh networks

Mesh networks can be utilised to increase the range of coverage, where the 'mesh' is an interconnection among a network of devices. Only one device in the mesh needs to be connected to the network, which can then relay data to other nearby devices.

Mesh networks not only provide the opportunity to support 5G deployment but also have the additional benefit of providing efficient network speeds. Through dynamic routing, devices on a mesh network are able to seek the fastest and most reliable pathway to send and receive data. As such, this architecture can provide a cost effective solution for coverage in more remote areas. For example,

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 6 of 12

primary industries would benefit from this network approach by simplifying network connections and costs associated with deploying and managing an IoT sensor network. However, the application of mesh networks is still highly speculative as the industry considers how they will operate in practice and in different environments. This may contribute to the development of new business models for 5G.

Spectrum sharing

Spectrum is a critical enabler of Australia's current and future communications infrastructure. The specific spectrum bands and quantity of spectrum required for 5G are still being considered. However, 5G will likely require a mix of low, medium and high frequency spectrum to meet different scenarios relating to coverage, connectivity, and latency:

- Low frequency (less than 1GHz)—providing widespread coverage across urban, suburban and rural areas and supporting IoT for low data rate applications.
- Medium frequency (1–6GHz)—providing good coverage and high speeds, and including the expected initial 5G range of 3.3–3.8GHz which has been identified as the most likely band for launching 5G globally.
- High frequency (above 6GHz)—providing ultra-high broadband speeds for advanced mobile broadband applications, and most suitable for applications in dense traffic hotspots.

5G technologies can be expected to deliver improvements in spectral efficiency (the data rate that can be supported per unit of spectrum). However, the use of 5G networks for applications such as widespread industrial applications is likely to require significantly more amounts of contiguous spectrum to be made available.

While some of the potential bands for 5G currently have unused spectrum, other bands would need to be 'refarmed', noting that it is likely that bands currently used for 2G, 3G and 4G in Australia will transition over time to 5G. Refarming enables spectrum to be transitioned to the highest value use as required. The refarming of spectrum already held by mobile broadband operators is a commercial decision for those operators. 5G is also expected to provide the opportunity for 'soft-refarming' where 4G and 5G technologies can both be supported simultaneously, minimising the impact to legacy devices during transition periods.

Spectrum sharing, that is spectrum accessed by numerous users on a shared basis, has also been identified as an option for 5G technology. Spectrum sharing encompasses a range of different aspects of spectrum management. Spectrum can be shared by geography, time, economic priority schemes, code modulation, polarisation, directionality or power. Access to spectrum is divided between users so it can be used without interference issues.

5G opens up new opportunities for increased spectrum sharing, through mechanisms such as network slicing. 5G technology is also designed to support shared arrangements, and allows for the sharing of the same spectrum ('unlicensed Wi-Fi' spectrum) with other technologies. Operators can augment their holdings in situations where existing exclusive holdings are insufficient to meet customer needs. Spectrum sharing in the 5G context is also supported by the expected use of highly directional antenna technology which would enable operators to operate in closer proximity without interference.

Antenna technology and network topology

5G will require radically different structures of networks if it is to achieve successful deployment in Australia. As 5G will likely utilise different frequencies, new equipment will be necessary. Additionally, the higher frequency 5G spectrum can only travel a small distance and will need more cells to ensure adequate coverage. However, antennas and equipment will be smaller, making it easier to attach these cells to existing infrastructure such as street lights and buildings.

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 7 of 12

The more dense deployment of cells will also give rise to other approaches that improve the reliability of data transmission across a 5G network. Data may be divided into individual streams and transmitted through multiple antenna segments in a process called Multiple Input Multiple Output (MIMO) which allows for more information to be transmitted simultaneously. This technique is further empowered by 'beamforming' which allows base stations to direct focused beams of energy to a specific area rather than dissipate the available power of a larger area. These developments will enable more efficient transmission and increase overall throughput.

5G is also expected to assist the adoption of IoT by further reducing power consumption through extended discontinuous reception. In this scenario, IoT devices shift between active and inactive cycles, transmitting only when required. This will allow connected devices to operate for extended periods on a single charge, reducing operational costs.

5G will also increase the support for a greater density of devices that would have otherwise been limited by the capacity on 3G and 4G networks. They will also enhance IoT deployment through the use of network slicing to create virtual network configurations that are optimised for the low power and coverage requirements of IoT networks.

Business case

For Australians to experience the benefits of 5G, the communications sector will need to explore and develop new business cases to attract investment and support the rollout of 5G services. As 5G becomes an integral part of the communications ecosystem, the sector will need to be agile to respond to the needs and expectations of other sectors which will be seeking to take advantage of these next generation networks.

While residential consumers will inevitably be attracted to the enhanced mobile broadband services offered by 5G, it is the industrial applications for 5G where industry expects to see the greatest opportunity for new business models.

Compelling industrial applications are still to be developed. For example, industry sectors can be expected to seek tailored solutions for their business needs such as enhancing their local area network or enabling autonomous systems, and small businesses will look for low cost deployments of IoT.

It is therefore expected that the model for 5G will be demand driven and will require the communications industry to foster new business opportunities.

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 8 of 12

How are countries preparing for 5G?

In preparation for 5G, many countries have been taking steps to test the technology and review their spectrum arrangements.

Examples of 5G work being undertaken by other countries

The United States is clearing the 600MHz band through an incentive auction for the potential early deployment of 5G. The United States has also identified reforms to infrastructure deployment as a priority for 5G rollout. Mobile network operators are also conducting trials of low-band spectrum for use in 5G services.

The Asia-Pacific Telecommunity (APT), the regional arm of the ITU, approved a recommendation for the use of the 700MHz band for 5G, with 26 countries in the Asia-Pacific region identifying this band for this use, including Australia, Japan, South Korea and New Zealand.

Korea and Japan have stated their intention to use some or all of the 26.5 to 29.5GHz range to trial enhanced mobile broadband applications ahead of the 2018 Winter and 2020 Summer Olympics respectively. It is expected that these trials will lead to commercial availability of 5G services.

The European Communications Commission (ECC) identified the 3.4–3.8 GHz and 26GHz band (from 24.25 to 27.5GHz) bands for the deployment of 5G in Europe. All European countries are expected to select at least part of this range to launch 5G by 2020.

In the United Kingdom, the government launched its 5G strategy in March 2017. It is funding testbeds to understand the different deployment requirements and security considerations for 5G. The United Kingdom is also working to make suitable spectrum available in the high (24.25 GHz–27.5 GHz, and other bands above 30 GHz), medium (3.4–3.8 GHz) and low frequency (700 MHz) bands.

Industry 5G preparations in Australia

Australia is well positioned to harness the opportunities of 5G. Australia has an effective and competitive mobile communications market, with voice and data coverage available to more than 99 per cent of the population.[4] It is the top performer internationally in terms of having in place effective enablers—infrastructure, affordability, consumer readiness and content availability—to support mobile internet adoption.[5]

5G trials have already commenced in Australia, with each of the main carriers working with mobile equipment suppliers in testing the application and limits of the technology. These trials will continue and will inform the communications sector on how 5G can be effectively deployed for the Australian environment.

---

[4] Department of Communications and the Arts, accessed 1 September 2017, 'Mobile phone towers'.
[5] GSM Association, 24 June 2016, 'Global Mobile Connectivity Index'.

www.communications.gov.au
5G—Enabling the future economy   www.arts.gov.au   Page 9 of 12
www.classification.gov.au

The Government's direction for 5G

The Government has made significant investments to improve telecommunications infrastructure in Australia, through the NBN and the Mobile Black Spot Program. The Government is also working to create a policy and regulatory environment that supports a competitive and innovative communications market.

The Government recognises that 5G will enable innovation and productivity across industry sectors and can significantly contribute to Australian's growth and future prosperity. Therefore, the Government will focus on enabling the early deployment of this new generation of mobile networks in Australia and encourage its use in delivering new services and applications.

The Government will support the timely rollout of 5G in Australia to enable the next wave of broad-based industry productivity, and support the growth of Australia's digital economy.

Industry expects and needs to lead the deployment of 5G. However, the Government has a role in supporting network rollout by modernising policy and regulatory frameworks and removing barriers that would delay rollout and adoption unnecessarily.

In the first instance, the Government will support the early deployment of 5G in Australia by:

- making spectrum available in a timely manner
- actively engaging in international spectrum harmonisation activities
- streamlining arrangements to allow mobile carriers to deploy infrastructure more quickly, and
- reviewing existing telecommunications regulatory arrangements to ensure they are fit-for-purpose.

Making spectrum available in a timely manner

A clear, efficient and flexible regulatory framework governing spectrum access will be essential to support the timely deployment of 5G networks in Australia.

The Government is currently undertaking work to modernise Australia's spectrum management framework to ensure it remains fit-for-purpose. In May 2017, it outlined its proposed reforms to the framework which are designed to simplify and streamline the processes for spectrum allocation and provide a transparent, efficient and flexible spectrum management framework. This will be the most significant change to the Australian spectrum management framework in the last 25 years.

The reforms will remove barriers between licence types, and enable flexible licensing issue and allocation processes. This strategic approach will remove outdated processes and support the Australian Communications and Media Authority (the ACMA) to more effectively respond to market demands and new technologies, such as 5G. This will help Australia remain internationally competitive with a modern, innovative economy over the coming decades.

The Government will put in place its new spectrum management framework by 2019.

In addition, the ACMA will continue to work on making spectrum available for 5G. The ACMA has been investigating the use of 1.5GHz and 3.6GHz and high frequency mmWave bands in considering additional spectrum for mobile broadband services. The ACMA has decided to prioritise refarming of the 3.6GHz band over the 1.5GHz band, citing industry submissions noting this band is likely to be a pioneer band for early 5G deployments and the need to provide greater clarity and investment certainty for incumbents and potential new band entrants alike. The ACMA is currently engaging with industry on which parts of the 3.6GHz band should be reallocated and on what terms. This approach also follows international trends which have seen the 3.6GHz band commonly used for 5G trials.

The ACMA will work to bring 3.6GHz spectrum to auction in 2018.

5G—Enabling the future economy
www.communications.gov.au
www.arts.gov.au
www.classification.gov.au
Page 10 of 12

Actively engaging in international spectrum harmonisation activities

There is already significant work underway globally with several countries trialling 5G, but standards for this new generation technology are yet to be finalised. The formal, international process to define 5G is led by the ITU. The ITU's Working Party 5D is responsible for shaping the standard for "futuristic mobile technologies" to support International Mobile Telecommunications (IMT) for 2020 and beyond.[6] This process is known as IMT-2020.

Stakeholders such as regulatory and policy setting bodies, hardware manufacturers and governments of countries in which they are based will be seeking to influence the international dialogue. A key body is the industry-driven 3GPP which undertakes technology standardisation. Collectively, the ITU and 3GPP will drive spectrum harmonisation activities: the ITU, led by administrations, will focus on the spectrum requirements; and the 3GPP, led by industry, will concentrate on equipment and device standards.

Industry is well-placed to lead the standardisation process particularly given its role identifying the application of 5G technologies. However, there is also an important role for the Government in these processes. This is particularly the case in the harmonisation of international spectrum arrangements which will have significant impact on the availability and cost of 5G devices in Australia and can be strongly contested.

In the past, Australia and the Asia-Pacific region have been influential in contributing to standards and spectrum plans that have been adopted across the world. This ensured that Australia was able to adopt new technologies quickly and that the Australian market could take advantage of the economies of scale and have a greater choice of mobile handset equipment.

The Government will ensure strong participation by Australian in domestic and international discussions about 5G spectrum harmonisation. Our continued involvement provides the opportunity to contribute to this dialogue and secure outcomes that will benefit the adoption of 5G in Australia.

Streamlining arrangements to allow mobile carriers to deploy infrastructure more quickly

The design and deployment of mobile networks will be radically different from those of today. 5G is expected to require additional infrastructure in new forms, including smaller cells and more densely-located antennas, particularly in the use of high-band spectrum.

Carriers have specific powers and immunities relating to telecommunications infrastructure deployment and installation. These laws help carriers to rollout telecommunications infrastructure quickly in a nationally-uniform way, rather than having to follow state, territory and local government requirements. These laws have existed in their current form since 1997.

The Government has recently consulted on proposed new arrangements that take account of technology developments and changes in operating practices as well as identifying opportunities to streamline deployment processes. Under these new proposed arrangements, mobile carriers would be able to rollout new communications technologies such as 5G more efficiently.

The Government understands that some members of the community have expressed concerns about the impact on public amenity from increased telecommunications infrastructure. The Government considers that telecommunications providers should work with local communities to address concerns about their infrastructure plans and is encouraging industry to consider consultation requirements for future 5G networks.

The Government continues to work with stakeholders and will implement the first tranche of changes to carrier powers and immunities following the conclusion of the consultation process.

---

[6] ITU, accessed 1 September, ITU towards "IMT for 2020 and beyond."

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 11 of 12

Reviewing existing telecommunications regulatory arrangements to ensure they are fit-for-purpose

The pace of change in the communications sector is impacting on the effectiveness of the existing telecommunications regulatory framework. In response, the Government is progressively working to modernise the regulatory architecture to ensure that the regulatory and policy settings flexibly respond to the current and future needs of the communications sector.

In undertaking reform, the Government is cognisant that the communications regulatory framework will need to be sufficiently flexible to address the emergence of new technologies and business models.

5G deployment will benefit from the Government's reform efforts, including updates to the regulatory framework for telecommunications and radio spectrum and modernising the ACMA. In these cases, the Government has been revising regulation to shift it away from the traditional vertical telecommunications sectors to principles based, flexible arrangements. This will provide the versatility necessary to account for 5G developments and services.

The Government will continue to work with industry to modernise current telecommunications regulatory arrangements to ensure they encourage competition and innovation in the sector.

As with previous mobile networks, cybersecurity will be a critical consideration as 5G is deployed. Security will be even more so a challenge for 5G, as the reliability of communications will be pivotal in the technology's ability to deliver benefits, particularly in the case of critical communications.

Additionally, in providing the architecture for automation, 5G networks will trigger an ever-increasing volume of data. While the bulk of data will be machine-to-machine communications, users will want assurance that their personal information is protected. User consent will be an area of growing complexity due to the intersection between autonomous systems and the individual.

Industry has strong incentives to address cybersecurity risks in 5G's new types of network deployments and systems. However, **the Government will continue to assess cybersecurity and privacy issues as they evolve to ensure Australians have confidence in using 5G.**

Next steps

The Government recognises the opportunities presented by 5G for economy-wide transformation, creating productivity benefits in sectors such as transportation, health, manufacturing and agriculture. However, realising the benefits of 5G in sectors other than communications will need the right sectoral regulatory settings.

The Government will work to ensure that sectoral regulatory frameworks are updated to take advantage of 5G. The communications portfolio is well-placed to facilitate an ongoing strategic dialogue on 5G that will support sectors to identify and work towards unlocking the potential benefits of the technology. This dialogue would provide a starting point for greater engagement across Government and with industry and the community.

To that end, **the Government will establish a 5G working group that will bring together representatives from across Government and industry.** The working group will create a platform for this strategic dialogue with a mandate to seek out opportunities and emerging issues on 5G. This will provide better coverage across Government of the evolving policy and regulatory challenges associated with 5G.

5G—Enabling the future economy

www.communications.gov.au
www.arts.gov.au
www.classification.gov.au

Page 12 of 12

Electrical and Computer Engineering Faculty
Publications and Presentations

Department of Electrical and Computer
Engineering

1-1-2016

Energy Efficient Switched Parasitic Array Antenna for 5G Networks and IoT

Ahmed Kausar
Boise State University

Hani Mehrpouyan
Boise State University

Mathini Sellathurai
Boise State University

Rongrong Qian
Heriot Watt University

Shafaq Kausar
National University of Sciences and Technology

Energy Efficient Switched Parasitic Array Antenna for 5G Networks and IoT

Ahmed Kausar*, Hani Mehrpouyan *, Mathini Sellathurai[†], Rongrong Qian[†], and Shafaq Kausar[‡]

*Department of Electrical and Computer Engineering
Boise State University, Boise, USA
Email: ahmedkausar@boisestate.edu, hanimehrpouyan@boisestate.edu

[†] School of Engineering and Physical Sciences
Heriot Watt University, Edinburgh, U.K
Email: m.sellathurai@hw.ac.uk, rq31@hw.ac.uk

[‡]Department of Electrical Engineering
National University of Sciences and Technology, Islamabad, Pakistan
Email: shafaqkausar@nust.edu.pk

Abstract—This paper includes design and implementation result of an adaptive beam forming antenna for upcoming 5G and Internet of Things (IoT). Switched parasitic array antennas are low cost, small sized and compact circular array antennas that steer beam in a desired direction by variation in switching pattern of parasitic elements. The proposed antenna design has an active center element, which is surrounded by several symmetrically placed parasitic elements. The designed antenna has a gain of 8 dB and is capable of 360 degrees beam steering in steps of 60 degrees each. Simulations are validated with results of the fabricated antenna. Antenna beam is steered by controlling parasitic elements. Future application of Electronically Steerable Parasitic Array Radiator (ESPAR) antennas and switched parasitic array antennas in next generation communication networks and methods for reducing size of the antenna are also highlighted.

Index Terms—Internet of Things, 5G, circular array, reconfigurable, smart antenna, adaptive beamforming, Switched-Parasitic Array

I. Introduction

Smart antennas are capable of adaptive beam forming in accordance with the environment. Over the last three decades wireless communication has made rapid progress due to better modulation schemes and better error correction schemes. Wireless communication systems are replacing the wired ones. Antenna is an integral part of wireless devices. To meet the ever-increasing demand of higher data rate, smart antennas capable of adaptive beam forming are introduced as a potential solution to increasing data rate and improving signal-to-noise ratio (SNR) in upcoming 5G and Internet of Things (IoT) [1]. By using smart antennas Equivalent Isotropically Radiated power (EIRP) can be increased. EIRP denotes product of the transmit power (P_t) and transmit antenna gain (G_t), i.e.

$$EIRP = P_t \times G_t \qquad (1)$$

From (1), it is observed that by using a smart antenna G_t increases, which in turn leads to the increase in the EIRP. The beam of an antenna can be steered by means of mechanical movement of a directional antenna or by electronic beam steering. Electronic beam steering is resilient and provides more accurate approximation of the desired signal source. Moreover electronic beam steering antenna does not involve moving parts. Therefore, there is less wear and tear on the antenna as compared to mechanical beam steering. Generally electronic beam forming is achieved by using phase shifters coupled with an array of antenna elements. The phase of an individual phase shifter is varied such that there is net constructive interference in the intended direction of the antenna beam and destructive interference in all other directions [2]. Hence, electronic beam steering is preferred over mechanical beam steering.

Electronically Steerable Parasitic Array Radiator (ESPAR) antennas use mutual coupling between antenna elements to steer the beams instead of phase shifters. Design of parasitic array radiator antennas is discussed in [3], [4] in which there is a center active element surrounded by a circular array of passive elements, where each passive element is loaded with a specific reactance. ESPAR antennas are smaller in size as compared to phase array antennas because elements have to be placed at a quarter of wavelength or less distance from each other for effective mutual coupling. Hence, ESPAR antennas are cost effective and space efficient. In such antennas we have only one active element, therefore, only one feed is required. Consequently, ESPAR antennas have lower losses (due to single feed) as compared to phase array antennas [3].

Previously designed antenna Electronically Steerable Parasitic Raidator(ESPAR) antennas have all cylindrical element, including the feed element [6], [7]. In this paper we have proposed design of seven element ESPAR antenna with conical center(active element)[1]. Length of conical element is optimized using optimetrics in HFSS. Our proposed antenna has higher bandwidth as compared to previously designed ESPAR antennas. [6], [7]. In multiple-input-multiple-output (MIMO)

---

[1]This work was supported by National Science Foundation (NSF) on Enhancing Access to Radio Spectrum (EARS)
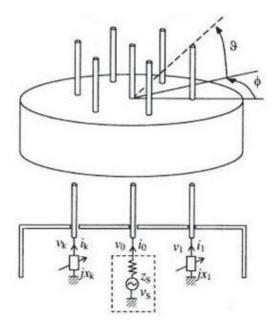
Fig. 1. Seven Element switched parasitic array antenna, ϕ is azimuth angle and θ is elevation angle.

Fig. 2. Z_{21} versus separation distance.

systems there are multiple transmit and and receive antennas. Each receive antenna receives signal over different propagation channels yielding diversity gain. The MIMO technique can increase system throughput and reduce propagation losses. Similarly diversity gain can also be achieved by using smart antennas. In particular, we can transmit the same signal in multiple directions through directional beams, such that they experience different propagation channels [4]. Our proposed antenna is more suited for MIMO applications compared to previously designed model because of its higher gain and improved bandwidth efficiency.

This paper focuses on the design of a seven element switched parasitic array antenna. The antenna is first simulated and then it is fabricated. Simulation results are validated with the hardware results. Previously done work on ESPAR and switched array antennas involve cylindrical monopole elements mounted on a ground plain [8] due to cylindrical nature bandwidth of active monopole is confined, subsequently bandwidth of switched array and ESPAR antennas is limited. Instead on cylindrical monopole element we have used tapered conical monopole. Due to tapered design bandwidth is enhanced.

II. SWITCHED PARASITIC ARRAY ANTENNA CONCEPT

The conceptual overview of the switched parasitic array antenna is given in Fig. 1. We have first simulated the antenna with seven monopole elements. There is one element at center which is surrounded by a circular array. Feed is applied to the center element and it is termed as active element, whereas elements in the circular array are parasitic elements. Parasitic array antennas do not involve phase shifters, so that the
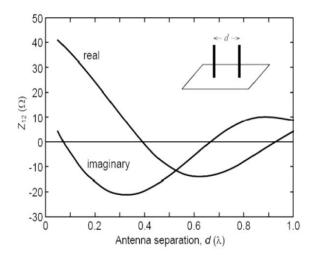
manufacturing cost is less as compared to the phase array antennas. Each element is separated from other element by a quarter of wavelength, thus mutual coupling phenomena comes into account and there is a shift in phase of current in each element because of this mutual coupling phenomenon. Since switched parasitic array antennas offer adaptive beamforming solution with compact size and low cost, they are likely to be used in future communication devices to increase data-rates [6], [7]. Parasitic array antennas have vast potential for usage in laptops, cell phones, WLANs etc.

A. Mutual Coupling

Mutual Coupling is basically the interaction between antenna elements in an array. Since parasitic elements are placed closer to each other (at a distance of $\lambda/4$) there will be mutual coupling.This mutual coupling phenomenon is responsible for a shift in phase of current induced in each element [8].The current induced by one antenna element in another antenna element [9] is given by (3)

$$I_2 = \frac{Z_{21}}{Z_{22}} \times I_0 \qquad (2)$$

$$Z_{21} = \left| \frac{V_2}{I_1} \right|_{I_2=0} \qquad (3)$$

$$Z_{21} = \left| \frac{V_2}{I_2} \right|_{I_1=0} \qquad (4)$$

Here Z_{22} is constant as it is open-circuit output impedance. Z_{21} is open-circuit transfer impedenace from port 2 to port 1 and it can be approximated by graph shown in Fig. 2. From equation (3), beam can be steered in the desired direction by varying the phase of current I_2 [10].
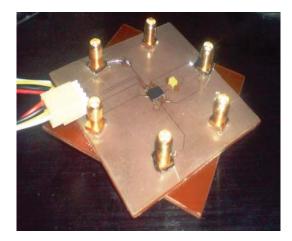
Fig. 3. RF PCB with Analog devices ADG 904 RF switches.

TABLE I
DESIGN PARAMETERS

Parameters	Values
Ground Skirt length	lambda/4
Monople Length	lambda/4
Frequency	2.45 GHz
Ground Radius	lambda/2
Monopole Radius	lambda/200
Ground Thickness	3 mm

B. Beam Steering Mechanism

In this work the designed antenna is capable of 360^0 beam steering in steps of 60^0. The beam steering is achieved by using two sets of 4:1 Analog devices multiplier switch ADG 904. Fig.3 shows RF PCB designed for beam steering. Beam is steered towards an element which is open. The open element behaves as a director and shorted elements act as reflectors. In order to achieve a maxima at 0^0, we make the element 1 open by using RF switch IC while other elements (i.e., elements 2,3,4,5 and 6) are shorted. This mechanism is similar to Yagi-Uda antenna where elements with shorter lengths in front of feed element act as directors and elements at rear of feed element acts as reflectors [11]. In Yagi-Uda antenna feed element is often a dipole or folded dipole.

III. PARASITIC ARRAY ANTENNA DESIGN

Design parameters and mechanical design of antenna is included in this section.

A. Design Parameters

Antenna design parameters are tabulated and the results are optimized using antenna simulation tools. Ground is skirted, since skirted ground provides mechanically sound model for PCB assembly to mount at back of antenna [12]. Skirt is $\lambda/4$ in length similarly each monopole element is lambda/4 in length and $\lambda/200$ is its radius.Table I shows the design parameters for the seven-element switched parasitic array antenna.

Fig. 4. Designed seven element parasitic switched array antenna.
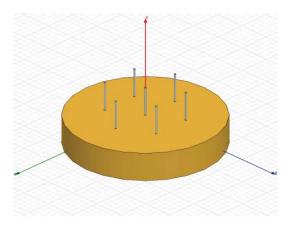
Fig. 5. Simulated antenna model.

B. Mechanical Design

Monopole elements are made by cutting copper wire of desired gauge i.e lambda/200 in radius [13] and these elements are soldered on pin of SMA connectors.Aluminum is used as ground plane. Electromagnetic waves are reflected by the ground. Instead of using seven dipole elements we have used seven monopole elements mounted on the ground plane [14]. Ground is made hollow in order to create space for mounting control circuitry [15]. Fig. 4 shows designed hardware.

IV. ANTENNA SIMULATION AND PLOTS

The antenna is first simulated and then simulated results are compared with the measured results of fabricated antenna. Simulated results were in line with the hardware results and 8 dB gain was observed in simulated results as well as in fabricated antenna. In simulation each of the parasitic element is loaded with lumped port excitation, simulating opening and shorting [16] of monopole elements. Fig. 5 shows simulated model.

TABLE II
MAXIMA PATTERN

Azimuth Angle	Element 1	Element 2	Element 3	Element 4	Element 5	Element 6
0 degree	Open	Short	Short	Short	Short	Short
30 degree	Open	Open	Short	Short	Short	Short
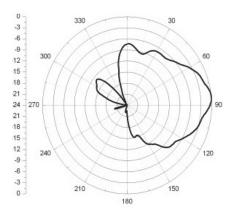
Fig. 6. Anechoic chamber results for $\theta=90^0$ (Beam in elevated because we are using monopole elements with finite ground).

Fig. 7. Anechoic chamber results for $\phi=0^0$.

Fig. 8. Anechoic chamber results for $\phi=30^0$.

Fig. 9. (a) Conventional ESPAR antenna model with cylindrical Center Element, Fig. 9 (b) Designed ESPAR antenna with conical element for Bandwidth Enhancement.

Two dimension polar plots depicts the beam steering in the direction of given elevation angle, theta (θ) and azimuth angle, phi (ϕ). Anechoic chamber results for $\theta=90$ and $\phi=0$ and 30 degrees are shown in Fig. 7 and 8 respectively. Using large ground plane angle of elevation is reduced closer to 90^0. Fig. 6 shows anechoic chamber result for $\theta=90$ one can observe small elevation form desired x-y axis because of ground skirt [17].In simulations elevation angle of 85^0 is observed. Fig. 6(a) shows antenna simulated plot for elevation angle θ and Fig. 6(b) shows anechoic chamber results.Anechoic chamber plot of Fig. 7 shows maxima at $\phi=0^0$, in order to get maxima at 0^0 elements from element number 2-6 are shorted and the element 1 is opened. Similarly Anechoic chamber plot of Fig. 8 shows maxima at $\phi=30^0$, in order to get maxima at 30^0 elements from element number 3-6 are shorted and element 1 and element 2 are open.

Traditional ESPAR antenna and switched parasiric array antenna have all cylindrical elements, including the active element [3], [4]. We have made center element conical instead of cylindrical. Due to conical nature of active element response our frequency range and improved. In order to validate the gain enhancement both models were simulated and results were compared. Fig 9. (a) shows conventional design and Fig. 9(b) shows simulated antenna with conical center element. Fig. 10 shows S_{11} plot for cylindrical active element. Considering -8_{dB} as cut off point bandwidth range from 2.358 GHz to 2.51 GHz (0.152 GHz). Fig. 11 shows $S-11$ plot for centre conical element, considering 8_{dB} as cut off point bandwidth ranges from 2.36 GHz to 2.555 GHz (0.195 GHz). Therefore my making center element conical we have improved bandwidth as compared to traditional switch array
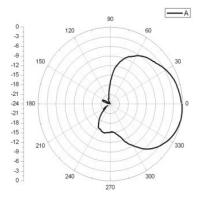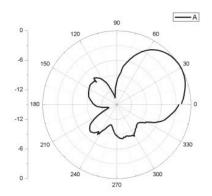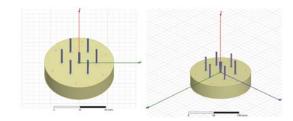
parasitic antennas. Low profile design for switched parasitic array antennas in also highlighted in [18].

V. FUTURE WORK AND CONCLUSION

5G networks will be designed for device to device, device to human and human to human interactions. High data rates are required to meet demands of 5G systems. One of proposed bands for 5G is in mm-Waves range i.e from 30GHz to 60GHz for such high frequencies path losses are considerable and beam steering antennas are way forward to overcome path
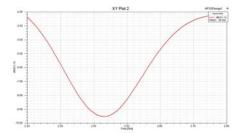
Fig. 10. S_{11} plot for conventional switched array antenna.
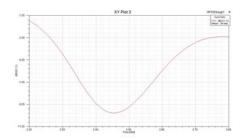
Fig. 11. S_{11} plot for center conical element switched array antenna.

losses [19]. ESPAR and switched parasitic array antennas provide a cost effective way of incorporating adaptive beamforming in future 5G and internet of Things (IoT) devices [20]. However for the efficient use of ESPAR antennas in mobile terminals the size of the antenna system must be reduced, for this further research on dielectric switched parasitic array antennas is underway. Seven element switched parasitic array antenna is designed that is capable of dynamic beam steering. Gain of 8dB is achieved practically. Simulated results are in line with the anechoic chamber results. Designed antenna increase the channel capacity by improving signal to noise ratio. Gain of 8dB was achieved in specified directions of elevation angle ,theta (θ) and azimuth angle, phi (ϕ). Antenna is optimized in terms of interference reduction for its use in wireless ad hoc networks, operating at frequency of 2.45 GHz. Bandwidth of designed antenna is 450 MHz and beam is steered in 360^0.

ACKNOWLEDGEMENT

REFERENCES

[1] M. Rzymowski, P. Woznica, and L. Kulas, "Single-anchor indoor localization using espar antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 1183–1186, 2016.

[2] L. Zhou, F. A. Khan, T. Ratnarajah, and C. B. Papadias, "Achieving arbitrary signals transmission using a single radio frequency chain," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4865–4878, Dec 2015.

[3] V. I. Barousis and C. B. Papadias, "Arbitrary precoding with single-fed parasitic arrays: Closed-form expressions and design guidelines," *IEEE Wireless Communications Letters*, vol. 3, no. 2, pp. 229–232, April 2014.

[4] M. R. Islam and M. Ali, "Elevation plane beam scanning of a novel parasitic array radiator antenna for 1900 mhz mobile handheld terminals," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 10, pp. 3344–3352, Oct 2010.

[5] S. Kausar, H. U. Rahman, T. Hassan, and A. Kausar, "Miniaturization of espar antenna using folded monopoles and conical central element," in *Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), 2015 International Conference on*. IEEE, 2015, pp. 87–91.

[6] H. T. Liu, S. Gao, and T. H. Loh, "Electrically small and low cost smart antenna for wireless communication," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 3, pp. 1540–1549, March 2012.

[7] T. H. Loh, H. Liu, and S. Gao, "A cost-effective direct magnitude measurement methodology for smart antennas," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 2043–2050, April 2013.

[8] F. Khan, K. Ali, A. Kausar, and S. Kausar, "Energy generation capacity analysis of a canal based hydro project," in *2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation*, Sept 2013, pp. 277–281.

[9] A. Kausar, M. A. Cheema, S. Kausar, H. Rehman, and T. Hassan, "Espar antenna system designing & simulation," in *Proceedings of the 2014 First International Conference on Systems Informatics, Modelling and Simulation*, ser. SIMS '14. IEEE Computer Society, 2014, pp. 193–196. [Online]. Available: http://dx.doi.org/10.1109/SIMS.2014.37

[10] S. Akkar and A. Gharsallah, "Reactance domains unitary music algorithms based on real-valued orthogonal decomposition for electronically steerable parasitic array radiator antennas," *IET Microwaves, Antennas Propagation*, vol. 6, no. 2, pp. 223–230, January 2012.

[11] S. Kausar, H. u. Rahman, A. Kausar, and T. Hassan, "Espar antenna system for dynamic tracking of active targets," in *Modelling Symposium (EMS), 2013 European*, Nov 2013, pp. 533–535.

[12] A. Umar Kausar, F. Aslam Khan, T. Hassan, A. Nasser, and S. Kausar, "Adaptive espar antenna system modelling & design." *International Journal of Simulation–Systems, Science & Technology*, vol. 14, no. 4, 2013.

[13] D. v. B. M. Trindade, C. Mller, M. C. F. D. Castro, and F. C. C. D. Castro, "Metamaterials applied to espar antenna for mutual coupling reduction," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 430–433, 2015.

[14] S. Kausar, Ahmed Kausar, "Smart adaptive beam forming antenna for interference cancellation," in *Wireless Days (WD), 2013 IFIP*. IEEE, 2013, pp. 1–3.

[15] A. Kausar, S. Kausar, and T. Hassan, "Smart adaptive beam forming antenna for interference minimization," in *Future Generation Communication Technology (FGCT), 2013 Second International Conference on*. IEEE, 2013, pp. 6–9.

[16] F. Aslam, E. Ulhaq, A. Nasser, and A. Umar, "Intelligent modeling scheme for detection of line losses in power distribution system," in *Computer Modelling and Simulation (UKSim), 2013 UKSim 15th International Conference on*. IEEE, 2013, pp. 218–223.

[17] R. Qian, M. Sellathurai, and D. Wilcox, "A study on mvdr beamforming applied to an espar antenna," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 67–70, Jan 2015.

[18] A. Pal, A. Mehta, D. Mirshekar-Syahkal, and H. Nakano, "A low-profile switched-beam dual-band capacitively coupled square loop antenna," in *Antennas and Propagation Conference (LAPC), 2013 Loughborough*, Nov 2013, pp. 563–566.

[19] A. Kausar, S. Kausar, T. Hassan *et al.*, "Interference cancellation using adaptive beam forming for 3 d beam steering." *International Journal of Simulation–Systems, Science & Technology*, vol. 14, no. 1, 2013.

[20] M. R. Nikkhah, P. Loghmannia, J. Rashed-Mohassel, and A. A. Kishk, "Theory of espar design with their implementation in large arrays," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 6, pp. 3359–3364, June 2014.

[21] T. Hassan, A. Kausar, H. Umair, and M. A. Anis, "Gain optimization of a seven element espar antenna using quasi-newton method," in *Microwave technology & computational electromagnetics (ICMTCE), 2011 IEEE international conference on*. IEEE, 2011, pp. 293–296.

Recent Antenna System Technologies for Next-generation Wireless Communications

Koichi Tsunekawa†

Abstract

This paper describes the requirements of the antennas or antenna systems in recent high-speed and high-capacity wireless communication systems and the key technologies for satisfying them. It also explains the antennas and antenna systems being researched and developed for next-generation wireless communication systems. The antenna in such a system requires high gain, high-efficiency antenna elements and a multi-antenna design. This paper introduces a high-efficiency antenna created by integrating it with a monolithic microwave integrated circuit (MMIC) in a millimeter-wave frequency-band system, a multi-antenna system for the multiple-input multiple-output (MIMO) technique that enables multiple data transmission for land mobile communication systems, and a beam-scanning antenna for satellite communication systems. It also explains the general research and development trend of antenna systems and technical problems in each wireless system as an introduction to the other selected papers in this issue.

1. Introduction

Wireless access technologies have advanced rapidly [1] and broadband wireless communication systems have become popular in our daily lives. An antenna is a very important component of a wireless system because it acts as the input and output interface for wireless equipment, as shown in **Fig. 1**. However, system performance is determined by the total performance of each component, such as the antenna and the radio-frequency (RF), intermediate-frequency (IF), and baseband (BB) units. For recent high-performance wireless systems, overall system design has been essential. Accordingly, the antenna must not
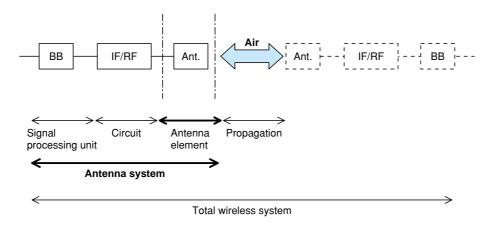
Fig. 1.   Wireless and antenna system.

† NTT Network Innovation Laboratories
   Yokosuka-shi, 239-0847 Japan
   E-mail: tsunekawa.koichi@lab.ntt.co.jp

only have antenna elements with high-performance characteristics, but must also perform effectively as an antenna system in conjunction with the RF, IF, and BB units. New functions can be added through antenna system design: for example, appropriate radiation pattern control provides a large system capacity by using baseband signal processing and an antenna element with an integrated RF circuit provides high gain without feeder loss.

The selected papers in this issue describe the latest antenna and antenna system technologies for achieving high performance. Two of the papers present beam-scanning antenna systems in satellite broadband communication. Beam scanning is carried out using by two methods: one [2] uses a variable active device or tunable reactance device incorporated into the antenna element, which has high-speed electronic control while the other [3] uses mechanically controlled antenna movement, which has a very precise direction control function. Another paper [4] introduces an active integrated antenna system with an RF circuit, which provides high gain and high efficiency, for a high-capacity millimeter-wave transmission system. And two papers describe a multiple-input

multiple-output (MIMO) antenna system for obtaining high-speed transmission and high system capacity in the land mobile communication system from the viewpoints of beamforming techniques [5] and implementation methods [6].

As an introduction to these advanced antenna techniques, this paper provides a general explanation of the requirements of antennas and antenna systems in recent high-speed and high-capacity wireless communication systems and the key technologies for satisfying these requirements. It also describes current antenna systems and NTT's antenna research results for typical wireless systems.

2. Antenna requirements in recent wireless communication systems

The trend of wireless systems is shown in **Fig. 2**. This indicates the typical system transmission speed and number of antennas needed to achieve a particular system performance for each operating frequency. Wireless communication systems are moving towards high transmission speeds and high capacity by using higher system frequencies to enable a wide frequency

RF-ID: radio-frequency identification
PDC: personal digital cellular
PHS: personal handy-phone system
IMT-2000: international mobile
    telecommunications 2000

HSDPA: high-speed download packet
    access
802.11x: family of IEEE standards
UWB: ultrawideband
FWA: fixed wireless access
4G: fourth generation

  : ubiquitous system
  : land mobile system
  : short-range system
  : WLAN system
  : fixed wireless system

Fig. 2.   Advances of terrestrial wireless communication systems and their relationship to antenna technologies.

band and using advanced system technology to achieve high quality and signal multiplexing. The antenna requirements for these system properties are:

- High-gain and high-efficiency technology
- Multi-antenna technology

Increases in the gain and efficiency of an antenna are offset by higher propagation loss at high frequencies, so the received signal level needs to be higher. High gain is an age-old problem in antenna research and development. Multi-antenna technology improves the receiving signal quality and provides multiplexed signal transmission by using an antenna system that includes RF circuits, BB units, and antenna elements.

2.1 High-gain and high-efficiency antenna

Theoretical and experimental antenna gain and efficiency characteristics are shown in **Fig. 3**. The dependences of gain and beamwidth on horn dimension $L(\lambda)$ in a relatively large antenna are explained in Fig. 3(a) [7]. When L is more than 0.5 times the wavelength, the antenna's efficiency is about 100%; a larger antenna provides higher gain. However, the high-gain radiation pattern leads to a narrow beamwidth. Therefore, a mobile terminal needs to use beam-scanning or beam-tracking techniques to follow the transmitter while the terminal moves. On the other hand, the efficiency decreases if the antenna size is less than 0.5 wavelengths. The bandwidth of a linear dipole antenna versus its efficiency for various antenna sizes (D: diameter) is shown in Fig. 3(b). This graph was estimated from theoretical limits [8] and many experimental results. The antenna efficiency becomes approximately –5 dB when antenna bandwidth of 10% is provided using a linear dipole antenna with a length of 0.3 wavelengths. This figure shows that the efficiency and bandwidth have a tradeoff relationship and that the efficiency drops greatly as the antenna size decreases. Therefore, some clever ideas for the antenna element structure are needed to overcome these weak points.

2.2 Multi-antenna system

A multi-antenna system has several antenna elements, but it does not merely work as an ordinary array antenna. The performance requirements are satisfied because the antenna system includes RF circuits, BB units, and antenna elements. The antenna element technologies and the multi-antenna system are shown in **Fig. 4**. In the antenna element, a high gain can be achieved by improvements to the antenna element structure, for example, by using a new material, adding a parasitic element, or integrating an RF circuit (Figs. 4(a)–(c)). The antenna shown in Fig. 4(c) is called an integrated antenna. It can control the radiation pattern as well as provide high gain by eliminating feeder loss. In a multi-antenna system (Fig. 4(d)), the radiation pattern is controlled by baseband signal processing, so the pattern of each wireless channel can be controlled by each baseband channel. Furthermore, it is possible to get multiplexed signal transmission in the space domain by baseband signal processing. Multi-antenna systems are generally called smart antennas or active antennas [9].

(a) Aperture antenna gain characteristics

(b) Dipole antenna efficiency characteristics
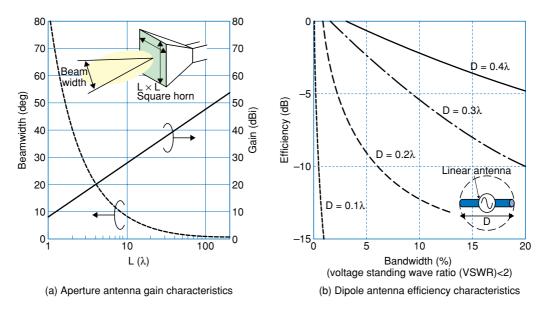
Fig. 3.   Antenna gain and efficiency characteristics.
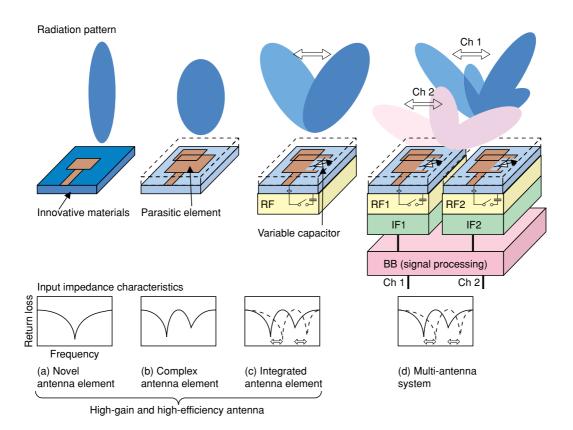
Fig. 4. Antenna elements and system.

An example of a mobile communication system using a smart antenna in the base station is shown in **Fig. 5**. Independent information can be sent to the terminals, which are spatially separated, using the same frequency at the same time because each beam is generated for a specific terminal by using the space division multiple access (SDMA) technique. Furthermore, N multiplexed transmissions can be sent to a terminal with N antennas by utilizing the slight difference in antenna locations in the terminal. MIMO is representative of this kind of technology. The latest smart antenna systems can increase the antenna gain even in a mobile system and expand the service area. The designed system capacity can be increased by using multiplexing techniques based on baseband signal processing. Multi-
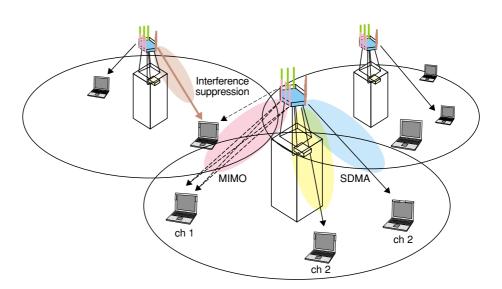
Fig. 5. Smart antenna system (land mobile telecommunication).

antenna systems can achieve very high performance with high-level functions corresponding to the wireless system requirements.

2.3 Key antenna technologies

The key technologies for achieving a high-gain,

high-efficiency, multi-antenna system are shown in **Table 1**. Antennas for the latest wireless communication systems are required to achieve high gain and high efficiency. The multi-antenna systems are designed to make maximum use of system resources, especially space.

3. Wireless systems and suitable antenna systems

The kinds of wireless systems that we can expect to find around us in the near future are illustrated in **Fig. 6**. RF-ID (radio-frequency identification) tags will be attached to various kinds of objects in a ubiquitous network and a large quantity of information will be gathered to provide useful knowledge to each person. High-speed data transmission ranging from a few megabits per second to 1 Gbit/s will be possible in mobile satellite systems, land mobile systems, and wireless local area network (WLAN) systems. In home networks, individual items of electronic equipment will be connected to each other at very high

Table 1.   Key technologies to meet antenna requirements.

Requirements	Key technologies
1) High gain and high efficiency	a. Innovative materials (meta-material, high magnetic material, …) b. Element design (parasitic element, fractal element, …) c. Integrated antenna (antenna with RF circuit or RF device, …) d. Active antenna (antenna within variable reactance or switch, ….)
2) Multi-antenna techniques	a. Beam scanning antenna b. Sector antenna (fixed narrow multi-beam system) c. Diversity antenna d. SDMA antenna e. MIMO antenna

Fig. 6.   Wireless communication systems.

Table 2.   Wireless systems and NTT's antenna system approaches.
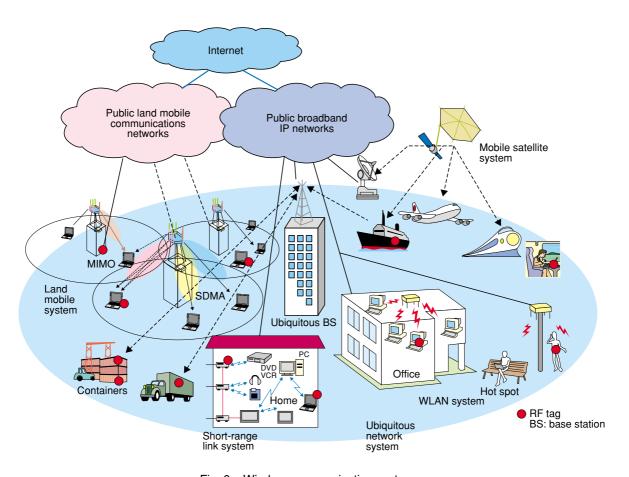
	System features	Main technical trends (example system, standard, or equipment)	NTT's antenna approaches (using technologies in Table 1)
Ubiquitous RF-ID	Multiple access Low rate	Active tags Sensor network Mesh network (802.15.5)	Compact high-efficiency tag antenna High-gain beam tilting base station antenna (1a, 1b, 1c)
Short-range communication	Short range Ultrahigh speed Link transmission	High frequency band (millimeter-wave system) Ultrawideband (UWB) Infrared rays (IrDA)	Active integrated antenna system (millimeter wave) (1b, 1c, 2a, 2b)
Land mobile and WLAN	Medium range High speed/capacity Many subscribers	MIMO-OFDM (802.11n,802.16a/e) MC-CDMA (4G, Beyond 3G) HSDPA (3.5G,1xEV-DO/V)	MIMO-OFDM antenna system Compact mobile terminal antenna system (1b, 1d, 2b, 2c, 2d, 2e)
Mobile satellite	Very long range Medium to high speed Various mobile stations	S/Ku band high-speed mobile comm. (N-Star/ETS-VIII) Broadband (Internet) (WINS) Millimeter wave/optical comm. (MILSTAR, GeoLITE)	Multi-beam antenna Large deployable antenna Beam-tracking antenna system (1c, 1d, 2a, 2e)

(discussed in papers in this issue)

speed (several gigabits per second) wirelessly instead of by cables. In these wireless systems, highly functional and high-performance antennas will be essential. **Table 2** explains the main features and technical trend of the systems and also shows NTT's approaches to antennas and antenna systems for each high-performance wireless system. The technologies in Table 1 that are used are indicated in parentheses. Each antenna system must satisfy severe specifications and also provide high functionality, as shown in Figs. 4(c) and (d). The antennas discussed in the selected papers in this issue are shown in blue in Table 2. The present status of each antenna system and research and development in NTT are explained below.

3.1   Ubiquitous network antenna system

Tag antenna elements have the same problems as pagers or portable telephone systems that use the LF, UHF, and VHF bands. The antenna element must be as small as possible and have high efficiency in a small volume and in the neighborhood of various objects. Most antennas for passive tags using the LF band (13 MHz) are spiral, helical, N-turn loop, or meander line antennas [10] for achieving an electromagnetic induction area and their gains are considered to be very low. Bent and spiral monopole or dipole antennas are used in personal computer card wireless communication systems or in small square film tags using UHF-band systems [11]. The approach is to design the antenna and RF-IC as one system; in other words, the connection impedance between the antenna and RF-IC is designed as a freely selectable value instead of being restricted to 50 Ω. The same kind of technique is currently being

investigated in NTT in a study of active integrated antennas [12]. On the other hand, the base station antenna can be designed by applying techniques developed for pagers or land mobile antennas because it has high-gain and beam-tilting characteristics.

3.2   Millimeter-wave frequency band communication antenna system

A high-speed wireless communication system is achieved by using wideband technology, so millimeter-wave frequencies, which provide a very wide bandwidth, are attracting attention. The use of high frequencies also requires high output power to compensate for the huge propagation loss and passive circuit losses, so high-gain and high-efficiency antennas are very important in system design [13]. Millimeter-wave systems mainly use horn, reflector [14], or lens antennas [15]. Horn or reflector antennas are used in relatively long-range communication where the range is several hundred meters. A lens type antenna, which is integrated with monolithic microwave integrated circuits (MMICs) to achieve high efficiency without a feeding line, is mainly used for short-range transmission, e.g., for a home link system. Active integrated antenna technology in which an amplifier for transmitting and receiving is mounted without using a connector or cables is suitable for a high-frequency-band system [16]. Typical antenna types for millimeter-wave frequency systems are shown in **Table 3**. Horn or reflector antennas provide very high gain but cannot be integrated with MMICs. The lens antenna provides high gain and can be made into an integrated antenna but requires adjustment to correct its

Table 3.   Antenna types for millimeter-wave systems.

Antenna	Gain	Integration with MMIC	Mounting adjustment
Horn or reflector	Very high (over 20 dBi)	No	(Feeder)
Lens	High (10 to 20 dBi)	Yes	Needed
Single patch	Low (2 dBi)	Yes	No
Parasitic element patch	High (10 dBi)	Yes	No
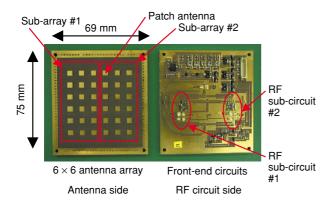
Fig. 7.   Active integrated antenna for 25-GHz system.

mounting error. The patch antenna is a good element for integration with MMICs because it can be made on the same material substrates, so it does not need antenna adjustment. We have developed an active integrated antenna designed for broadband mobile wireless access systems using the 25-GHz band, as shown in **Fig. 7** [12]. It exhibited output power of 14.6 dBm and a noise figure of less than 5 dB. The only weakness of this integrated antenna is that high gains cannot be provided with one patch antenna. The third selected paper in this issue describes a high-gain integrated active antenna and high-efficiency RF circuit for the millimeter-wave frequency band for a millimeter-wave communication system [4]. A high-gain integrated patch antenna is investigated as a parasitic element to obtain a wide antenna aperture. The millimeter-wave frequency-band link system was first proposed for application to the transmission of uncompressed HDTV (high-definition television) signals in the home or in a studio.

3.3   Land mobile and WLAN antenna systems

The number of subscribers to land mobile systems, especially cellular systems, has increased rapidly, and users are now demanding high-speed transmission

for broadband Internet access. To meets these demands, many improvements to the land mobile system have been proposed such as wideband code division multiple access (W-CDMA), adaptive modulation, and *ad hoc* transmission. Recently, the utilization of space has become as desirable as frequency and time utilization [17]. Space utilization is achieved using a multi-antenna system with a radiation control mechanism, which is accomplished by a smart antenna system.

3.3.1   Typical multi-antenna techniques for mobile systems

The application areas of multi-antenna techniques depend qualitatively on the system characteristics, as shown in **Fig. 8**. A fixed high-gain beam antenna does not always need antenna elements, but the other techniques are achieved through various combinations of antenna elements, so they are provided by adaptive or integrated multi-antenna systems, as shown in Figs. 4(c) and (d). A narrow-beam or high-gain antenna is used in the line-of-sight situation (where there are no obstacles between the transmitting and receiving antennas and there are few paths in the multipath transmission), so a beam-scanning mechanism is needed when a terminal moves fast, and multiple beams or sector beams are required when there are many terminals in the service area. The diversity antenna technique can provide a stable signal level in a multipath mobile environment. Beam scanning (with a sector beam or multiple beams) and diversity antenna techniques have been used fairly extensively from the early days of mobile communication service because they use only simple signal processing. However, with the rapid progress of signal processing technology, a desired signal can now be distinguished from a mixture of many signals in a multipath environment. The SDMA technique distributes signals among the terminals communicating in the service area. MIMO techniques enable very-high-data-rate communication with multiplexed transmission in each antenna of the multi-antenna system of the base station or mobile terminal. These techniques require more calculation as the movement speed of the mobile terminal increases. In these wireless systems, signal processing ability and superior algorithms are more important than the performance of antenna elements.

3.3.2   Research results and developed antenna systems

A very small and thin planar six-sector antenna using a patch Yagi-Uda array with common directors for mobile terminals has been proposed (**Fig. 9**) [18],
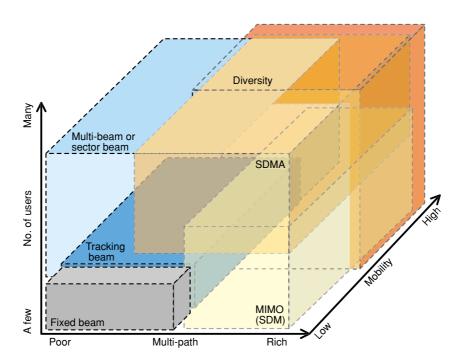
Fig. 8.   Dependence of multi-antenna techniques on system characteristics.
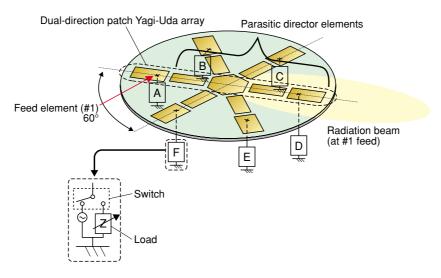
Fig. 9.   Thin and small sector beam antenna.

[19]. It has square parasitic director elements that are shared by two sector arrays that face in opposite directions, and it has a hexagonal central director element that is shared by all six sectors. These elements form three line arrays, which intersect at 60°. These shared elements enable the sector antenna to be compact. Experimental results showed that the fabricated antenna had a conical plane beam width of 63° and high front-back ratio of 17 dB. The creation of adaptive (smart) antenna hardware has been investigated [20]. A novel beam combination method for wide-

band digital indoor communication systems such as WLANs has been proposed [21] in NTT. Crossed fan beams are used at the base station and mobile terminal. The beam combination method provides better transmission quality than the traditional pencil-beam combination and eliminates the need for complex beam control in the base station. Maximum dispersion distributions in the transmission frequency bandwidth have been calculated using a three-dimensional indoor propagation delay simulation algorithm. A method of designing adaptive antennas including an

antenna configuration that is suitable for street microcells, considering the propagation environment, has been described. An automatic calibration method at the transmitters and receivers using the transmitting signal (ACT), which enables realtime calibration, has been presented [22]. An SDMA technique using polarization has also been studied [23]. The transmission quality of an SDMA scheme that uses vertical pattern and polarization control was investigated in an actual cellular environment. For a sector cell system, this configuration was shown to be effective based on an evaluation of the spatial correlation characteristics. We also found that the output carrier-to-interference-plus-noise power ratio (CINR) of the proposed configuration was from 10 to 17 dB higher than that of the conventional SDMA configuration when the number of users was greater than four. The proposed configuration can reduce the antenna size compared with conventional SDMA. Moreover, MIMO techniques enable very-high-data-rate transmission using a limited frequency band. We have studied technologies based on the eigenmode-space division multiplexing (E-SDM) [24] scheme because it provides the theoretically maximum channel capacity and reduces the terminal's calculation load. The selected papers in this issue introduce both theoretical and practical technologies. The fourth paper describes new MIMO methods with beamforming to overcome problems of calculation complexity and low MIMO effect in the line-of-sight scenario [5]. The fifth paper describes the configuration of an 8×4 MIMO-OFDM transceiver and presents experimental results for the propagation characteristics, bit error rate (BER), and frequency utilization [6].

3.4　Satellite broadband communication system

If we want to create a seamless and flexible wireless network system, then a satellite communication system is very important because it provides a wide service area. However, it has the inherent problem that the distance between the transmitter and receiver is long, so the terminal needs a larger antenna than ones in the land mobile system to provide high-speed transmission. The effective solution is to make the satellite dish larger to provide high gain and multibeam transmission. We have studied multibeam communication satellite technology and onboard large deployable antennas for creating a broadband mobile satellite communication system [25]. An MMIC-based beamforming network equipment for a multi-

Table 4.　Typical antennas in earth stations.

Satellite category	Typical satellite system	Current typical earth station antenna (size or length, gain)
LEO	Iridium Globalstar	Helical (0.1 m, 2 dBi) Linear whip (0.1 m, 0 dBi) Patch (0.1 m^2, 3 dBi)
MEO GEO	ICO N-STAR INMARSAT	Parabola (0.6 – 0.9 m, 18 – 21 dBi) Phased patch array (0.5 m^2, 14 dBi) Helical (0.3 m, 2 dBi) Patch array (0.3 m^2, 12 dBi) Patch (0.1 m^2, 3 dBi)

LEO: low earth orbit
MEO: medium earth orbit
GEO: geostationary earth orbit

beam feeder whose weight is only 1/3 that of the conventional type was developed. We are also developing a large, highly precise deployable antenna with a diameter in the over-10-m range. The mobile terminal acting as an earth station is also an important component of a broadband mobile satellite communication system. Typical existing antennas in earth stations are shown in **Table 4**. Various mobile terminals are accommodated in the satellite communication system by different transmission rates, and these terminals use various antenna methods and locations. A comparatively large terminal antenna is required to handle high data transmission rates in a satellite system to mitigate the huge propagation loss. The high gain and narrow beam of a large antenna make a beam-scanning mechanism necessary to enable the antenna to follow the satellite direction as the terminal moves. The sixth selected paper describes a highly accurate and cost-effective auto-tracking antenna system for earth stations onboard vessels, using a novel inclinometer that is highly accurate in acceleration disturbance environments and a new systematic stabilization controller design process based on H$_\infty$ control [3].

4.　Conclusion

This paper gave an overview of the requirements of antennas and antenna systems in recent high-speed and high-capacity wireless communication systems and the key technologies for satisfying these requirements to create next-generation wireless communications. It also described current antenna systems and NTT's antenna research results for typical wireless systems.

Antennas for next-generation wireless communications require high gain, high efficiency, and a multi-antenna design. A high-gain antenna can be made

using a relatively large antenna aperture, but a beam-scanning mechanism is needed for mobile communications. A high-efficiency antenna can be made using an active integrated antenna, which integrates an MMIC with the antenna element. Multi-antenna technology is the innovative technology that enables multiple data transmission in space. In MIMO-type multi-antenna systems, the most important key technologies are the beamforming algorithm and the implementation method. The solution technologies for NTT's advanced antenna systems are explained in detail in the following selected papers in this issue. The technological background and the current research and development situation are also described to make the papers easy to understand.

References

[1]   "Next-Generation Broadband Wireless Access Technologies," NTT Technical Review, Vol. 2, No. 1, pp. 12-63, 2004.

[2]   N. Honma, T. Seki, and K. Tsunekawa, "Planar Beam-Scanning Microstrip Antenna Using Tunable Reactance Devices for Satellite Communication Mobile Terminal," NTT Technical Review, Vol. 3, No. 9, pp. 24-32, 2005.

[3]   T. Yoshida, K. Ohata, and M. Ueba, "Highly Accurate and Cost-effective Auto-tracking Antenna System for Satellite Broadband Communication on Vessels," NTT Technical Review, Vol. 3, No. 9, pp. 60-71, 2005.

[4]   T. Seki, K. Nishikawa, I. Toyoda, and K. Tsunekawa, "Millimeter-wave High-Efficiency Multilayer Parasitic Microstrip Antenna Array for System-on-package," NTT Technical Review, Vol. 3, No. 9, pp. 33-41, 2005.

[5]   Y. Takatori, R. Kudo, K. Nishimori, and K. Tsunekawa, "Beamforming Method for Broadband MIMO-OFDM Systems," NTT Technical Review, Vol. 3, No. 9, pp. 42-49, 2005.

[6]   K. Nishimori, R. Kudo, Y. Takatori, and K. Tsunekawa, "Experimental Evaluation of 8×4 Eigenmode SDM Transmission in Broadband MIMO-OFDM Systems," NTT Technical Review, Vol. 3, No. 9, pp. 50-59, 2005.

[7]   J. D. Kraus, "Antennas, Sec. 2: Basic Antenna Concepts," McGraw-Hill Book Company, 1988.

[8]   R. C. Hansen, "Fundamental limitations in antennas," Proceedings of the IEEE, Vol. 69, Issue 2, Feb. 1981.

[9]   A. Alexiou and M. Haardt, "Communications smart antenna technologies for future wireless systems: trends and challenges," Magazine, IEEE, Vol. 42, Issue 9, pp. 90-97, Sep. 2004.

[10]   S. S. Basat, K. Lim, J. Laskar, and M. M. Tentzeris, "Design and Modeling of Embedded 13.56 MHz Rfid Antennas," IEEE AP-S Int. Antennas Propagat. Symp. Dig.,Vol. 4, Session 130, pp. 64-67, 2005.

[11]   U. Karthaus and M. Fischer, "Fully integrated passive UHF RFID transponder IC with 16.7-/spl mu/W minimum RF input power," Solid-State Circuits, IEEE Journal of Vol. 38, Issue 10, pp. 1602-1608, Oct. 2003.

[12]   T. Seki, F. Nuno, K. Nishikawa, and K. Cho, "Active Integrated Antenna Technique," NTT Technical Review, Vol. 2, No. 1, pp. 36-43, 2004.

[13]   T. Ihara and K. Fujimura, "Research and development of millimeter-wave short-range application systems," Trans., IEICE, Vol. E79-B, No. 12, pp. 1741-1753, Dec. 1996.

[14]   W. Menzel, D. Pilz, and M. Al-Tikriti, "Millimeter-wave folded reflector antennas with high gain, low loss, and low profile," IEEE Antennas and Propagation magazine, pp. 24-29, June 2002.

[15]   M. J. M. van der Vorst, P. J. I. de Maagt, A. Neto, R. M. Heeres, W. Luinge, and M. H. A. J. Herben, "Effect of internal reflections on the radiation properties and input impedance of integrated lens antennas-comparison between theory and measurements," Microwave Theory and Techniques, IEEE Trans. on 49, Issue 6, Part 1, pp. 1118-1125, June 2001.

[16]   T. Seki, K. Uehara, and K. Kagoshima, "A Three-Dimensional Active Antenna for a High-Speed Wireless Communication Application," IEEE MTT-S Int. Microwave Symp., Denver, USA, Vol. 2, pp. 975-978, June 1997.

[17]   G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," IEEE Journal on Vol. 17, Issue 11, pp. 1841-1852, Nov. 1999.

[18]   N. Honma, F. Kira, T. Maruyama, K. Cho, and H. Mizuno, "Compact six-sector antenna employing patch Yagi-Uda array with common director," Antennas and Propagation Society International Symposium 2002, IEEE, Vol. 1, pp. 26-29, June 2002.

[19]   N. Honma, T. Seki, K. Tsunekawa, F. Kira, and K. Cho, "Compact Multi-Sector Antenna Employing Microstrip Yagi-Uda Array Antenna with Common Director Elements," IEICE Trans. B, Vol. J87-B, No. 9, pp. 1363-1371, Sep. 2004 (in Japanese).

[20]   T. Hori, K. Cho, Y. Takatori, and K. Nishimori, "Key techniques for actualizing smart antenna hardware," Antennas and Propagation Society International Symposium 2000, IEEE, Vol. 2, pp. 948-952, July 2000.

[21]   K. Uehara, Y. Takatori, and K. Kagoshima, "Proposal for a novel beam combination method for indoor high-speed wireless communication systems," Vehicular Technology, IEEE Transactions on Vol. 48, Issue 5, pp. 1511-1517, Sep. 1999.

[22]   K. Nishimori, K. Cho, Y. Takatori, and T. Hori, "A Novel Configuration for Realizing Automatic Calibration of Adaptive Array Using Dispersed SPDT Switches for TDD Systems," IEICE Vol. E84-B, No. 9, pp. 2516-2522, Sep. 2001.

[23]   K. Nishimori, and K. Cho, "Evaluation of SDMA employing vertical pattern and polarization control in actual cellular environment measurement," Vehicular Technology Conference 2004, VTC 2004-Spring, 2004 IEEE 59th Vol. 1, pp. 244-248, May 2004.

[24]   A. Paulraj, R. Nabar, and D. Gore, "Introduction to space-time wireless communications," Cambridge, U.K., Cambridge Univ. Press, 2003.

[25]   M. Ueba, K. Ohata, J. Mitsugi, and M. Umehira, "Broadband and scalable mobile satellite communication system for future access networks," 22nd AIAA ICSSC, AIAA 2004-3154.

Koichi Tsunekawa

Senior Manager, Executive Research Engineer, Wireless Systems Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.S., M.S., and Ph.D. degrees in engineering science from the University of Tsukuba, Tsukuba, Ibaraki in 1981, 1983, and 1992, respectively. He joined the Electrical Communications Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT), Tokyo in 1983. Since 1984, he has been engaged in R&D of portable telephone antennas in land mobile communication systems. From 1993 to 2003, he was with NTT DoCoMo, Inc. where he worked on radio propagation research, intelligent antenna systems for wireless communications, and the development of IMT-2000 antenna systems. His research interests are antenna systems for MIMO transmission, millimeter-wave broadband access, and ubiquitous wireless system. He is a member of IEEE and IEICE.

Scientific
Research
Publishing

The Roles of 5G Mobile Broadband in the Development of IoT, Big Data, Cloud and SDN

Bao-Shuh Paul Lin[1,2], Fuchun Joseph Lin[1,2], Li-Ping Tung[1]

[1]Microelectronics & Information Research Center, National Chiao Tung University, Taiwan
[2]College of Computer Science, National Chiao Tung University, Taiwan
Email: bplin@mail.nctu.edu.tw, fjlin@nctu.edu.tw, lptung@nctu.edu.tw

Abstract

The fast technology development of 5G mobile broadband (5G), Internet of Things (IoT), Big Data Analytics (Big Data), Cloud Computing (Cloud) and Software Defined Networks (SDN) has made those technologies one after another and created strong interdependence among one another. For example, IoT applications that generate small data with large volume and fast velocity will need 5G with characteristics of high data rate and low latency to transmit such data faster and cheaper. On the other hand, those data also need Cloud to process and to store and furthermore, SDN to provide scalable network infrastructure to transport this large volume of data in an optimal way. This article explores the technical relationships among the development of IoT, Big Data, Cloud, and SDN in the coming 5G era and illustrates several ongoing programs and applications at National Chiao Tung University that are based on the converging of those technologies.

Keywords

1. Introduction

The roles of SDN, Cloud, IoT, and Big Data in 5G Networks have raised great interest recently [1]. According to the IEEE Computer Society (IEEE CS 2022 Report [2]), Cloud Computing (Cloud), Big Data Analytics (Big Data), Internet of Things (IoT), and Software Defined Networks (SDN) are among 4 of 20+ emerging technologies as illustrated in **Table 1**. In the meantime, based on the report in IEEE Communications Society (IEEE ComSoc Technology News [3]), among top 10 trends in 2015 as listed in **Table 2**, 5G, Virtualization (SDN & NFV), everywhere connectivity for IoT & IoE, and Big Data are also included. Combining both reports, we can identify 5G, Cloud, IoT/IoE, Big Data, and SDN as the five most worthwhile ICTs (information & communica-

Table 1. IEEE computer society 2022 report.

Technology & Indexing Terms		
1. Security Cross-Cutting Issues	9. Multicore	17. 3D Printing
2. Open Intellectual Property Movement	10. Photonics	18. Big Data and Analytics
3. Sustainability	11. Networking and Interconnectivity	19. Machine Learning and Intelligent Systems
4. Massively Online Open Courses	12. Software Defined Networks	20. Life Sciences
5. Quantum Computing	13. High Performance Computing	21. Computational Biology and Bioinformatics
6. Device and Nanotechnology	14. Cloud Computing	22. Robotics
7. 3D Integrated Circuits	15. Internet of Things	
8. Universal Memory	16. Natural User Interfaces	

Table 2. IEEE communications society top 10 trends 2015.

1)	5G
2)	FIBER EVERYWHERE
3)	VIRTUALIZATION, SDN & NFV
4)	EVERYWHERE CONNECTIVITY FOR IoT & IoE
5)	BIG DATA, COGNITIVE NETWORKS
6)	CYBERSECURITY
7)	GREEN COMMUNICATIONS
8)	SMARTER SMARTPHONES, CONNECTED SENSORS
9)	NETWORK NEUTRALITY, INTERNET GOVERNANCE
10)	MOLECULAR COMMUNICATIONS
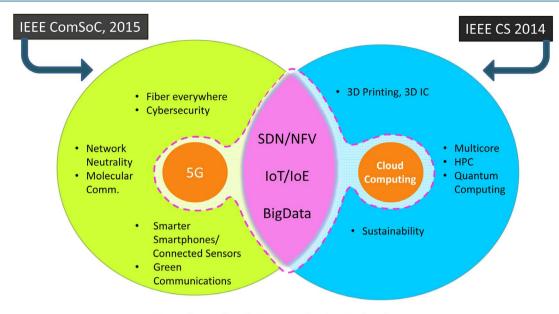
tions technologies) to watch out up to 2020 in term of their potential, convergence and applications. **Figure 1** depicts the overlapping of these five major ICT trends. It also illustrates how the infrastructure deployment of 5G mobile broadband and the architectural integration of Cloud Computing strongly impact the development of IoT, Big Data, and SDN. This paper explores the technical relationships of these five areas and discusses several ongoing programs and applications currently under development at NCTU based on these technologies.

2. The Roles of 5G, IoT, Big Data, Cloud, and SDN till 2020

Although so far the 5G mobile broadband requirements and standard specifications are not ready yet, 5G technology research & development are already started and some 5G features or subsystems are readily available. By Year 2020 the commercial 5G will be available and IoT applications will be deployed everywhere with mobile broadband technology. Moreover, the Big Data generated by IoT applications will become a norm and Cloud will be largely utilized to compute, store and virtualize network functions (NFV). Also, the underlying network infrastructure will adopt SDN to reduce both capital expense (CAPEX) and operational expense (OPEX). **Figure 2** further illustrates the roles of 5G, IoT, Big Data, Cloud and SDN and their relationships. This figure is modified based on the reference [4] by inserting 5G mobile broadband in the center.

3. Technical Relationships among IoT, Big Data, Cloud, & SDN in 5G Era

Based on **Figure 2**, we develop **Figure 3** to better explain the technical relationships among IoT, Big Data, Cloud, and SDN in the 5G mobile broadband services (5G MBS) [5]. First, IoT is capable of generating Big Data with four Vs: volume, velocity, variety and veracity. Then, Cloud is brought in for Big Data storage and processing. Finally, SDN is employed to provide more efficient and flexible networks for inter-Cloud data transport. Out of Big Data, Cloud, and SDN, advanced technologies such as machine learning analytics, Cloud

Figure 1. ICT major trends for 2015-2020.

Figure 2. Relationship among 5G, IoT, Big Data, Cloud and SDN based on [4].

RAN and softwarized 5G then are developed.

As illustrated in **Figure 4**, 5G will serve as a better gateway and transport network for IoT applications so that IoT data can be delivered more efficiently and economically. In addition, IoT will become one of the major sources of Big Data by producing large volume, fast velocity, and many varieties of data [6] as illustrated in **Figure 5**.

Finally, **Figure 6** shows that Cloud can be adopted in the 5G Radio Access Network (RAN) and turns it to a Cloud-based RAN (C-RAN). Both SDN and NFV have been applied to data center in the cloud to enable better load balance and resource allocation of the cloud. SDN has also been applied to 5G mobile broadband core networks to enable smart routing, better traffic management and improve network resource utilization.
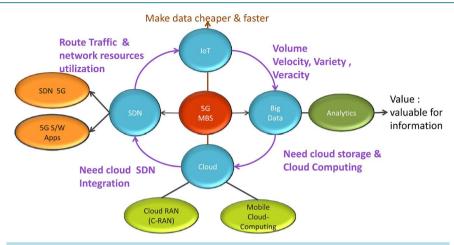
Figure 3. Technical relationships among 5G & IoT, Big Data, Cloud, and SDN.
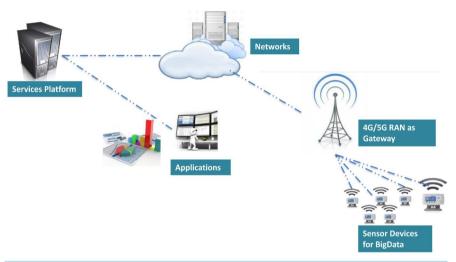
Figure 4. 4G/5G as the gateway for IoT applications.
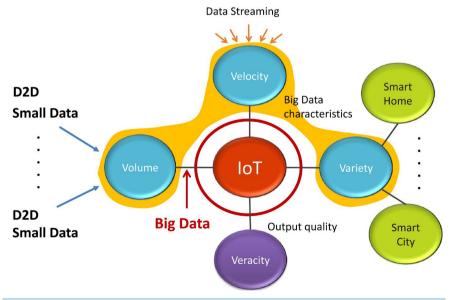
Figure 5. Relationships between IoT and Big Data.
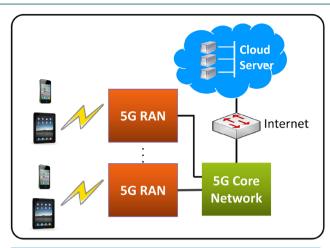
Figure 6. C-RAN architecture in 5G mobile network.

4. Ongoing Programs and Applications at National Chiao Tung University (NCTU)

In this section, we describe five ongoing programs and applications at NCTU that are closely related to the SDN, Cloud, Big Data, IoT, and 5G technologies [7]-[17].

4.1. Program on SDN-Enabled Cloud-Based Broadband and Wireless Network Technologies and Services

This program is a national program to support the SDN Industry-Academia Cooperation, led by both NCTU and CHT (Chunghwa Telecom, the largest mobile & telecom operator in Taiwan) with 5 other networking and communications companies in Taiwan. This is a multi-year program with the target to set up a testbed for end-to-end testing and to help establish an ecosystem for local SDN industry.

Figure 7 shows the scope of this program that covers mobile access, WiFi access, broadband core and data center cloud end-to-end application systems.

Figure 8 further shows both the structure of the program and its network configuration across NCTU, CHT and another university National Tsing Hwa University (NTHU).The program includes research and development involving 4G/LTE, B4G/5G, SDN, Cloud, SDN for Wi-Fi, and SDN for WAN technologies. There are five sub-programs: 1) SDN broadband network technologies and services, 2) SDN mobile/wireless network technologies and services, 3) SDN and cloud integration services and management, 4) SDN switch-related devices and systems and 5) SDN system integration and field trials.

The program is intended to construct an SDN network in NCTU, CHT and NTHU, respectively. These SDN networks then will be interconnected to form a wide area SDN network in Taiwan. Eventually, we plan to have this end-to-end system connected to the global network. The SDN switches deployed in this testbed will be mostly small or medium scale due to their experimental nature. Also included in this testbed are SDN-based Wi-Fi access points. Cloud-based data centers will be deployed in each location of NCTU, CHT and NTHU. Three types of SDN controllers including OpenDaylight, Ryu and Floodlight are currently under trial in the experimental network. Our architecture assumes the scenario of multiple SDN controllers with a hierarchical topology on which useful SDN APPs such as load balancer, network optimizer for video delivery, network co-ordinator (including visualization), dynamic flow configuration, end-to-end service configuration, dynamic policy-based traffic engineering, and multitenant network automation can be developed and deployed.

4.2. Program on Big Data Analytics for Network Traffic and Management Data

In this program, we address the network performance issues with two experimental networks: BML at NCTU campus and ITRINET of ITRI.

The architecture of the BML experimental network is illustrated in the lower part of Figure 9 that consists of a 4G RAN, a 4G Core and a Cloud environment. For the 4G RAN, both indoor and outdoor environment are taken into consideration. The left upper part of Figure 9 shows the scope of the ITRINET experimental network
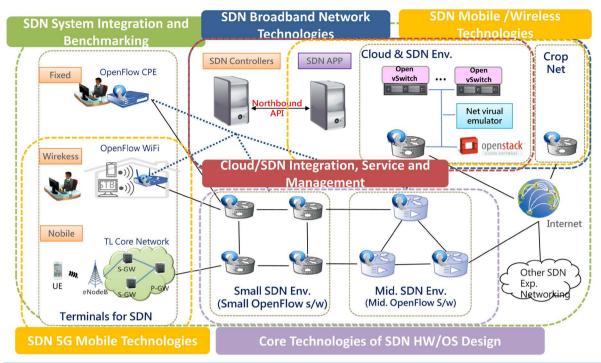
Figure 7. Applications of SDN & cloud in 4G/LTE & 5G wireless networks.
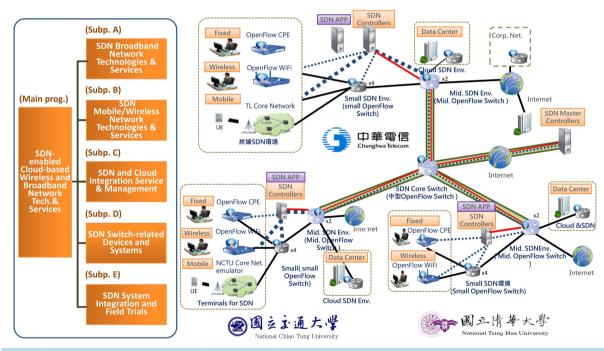
Figure 8. SDN technical R&D center and projects at NCTU.

that covers a much greater area of Hsinchu County than NCTU and includes additionally ITRI (Industrial Technologies Research Institute), HSIP (Hsinchu Science Industrial Park) and THR (Taiwan High-speed Rail, Hsinchu station). The right upper part of **Figure 9** illustrates how ITRINET covers various R&D buildings in the corporate campus of ITRI such as B11, B12, B51, etc.

Figure 10 illustrates how Big Data analytics based on InfoSphere or Spark is performed on ITRINET for the purpose of network optimization. First, network traffic measurement and network management data are collected.
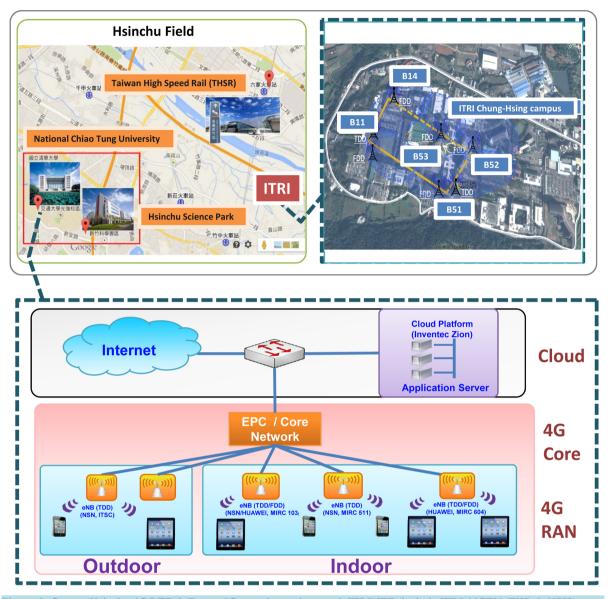
Figure 9. Greater Hsinchu 4G/LTE & Future 5G experimental network ITRINET (include ITRI, NCTU, THR & HSIP).

Then, data analytics methods based on machine learning, data mining and statistical modeling are applied to analyze the collected data. Finally, we apply the results thus generated to network performance evaluation and optimization by providing a feedback loop for system re-configuration. The whole operational cycle includes the technologies of 4G/LTE, B4G/5G, Big Data Analytics, Cloud, and Traffic Engineering.

4.3. Application on IoT Platform Integrated with Data Generation and Data Analytics

This is an application where we set up an IoT platform integrated with data generator and data analytics capabilities as illustrated in **Figure 11**. A common challenge for IoT/M2M service providers is how to test their large scale IoT/M2M applications with the near realistic data that the system will handle in a production environment. As such tests may involve not only a large number but also a large variety of sensors, deploying a testing environment that contains all the necessary sensors turns out to be an infeasible, if not impossible job. To tackle this problem, we develop a data generation method (illustrated in the lower part of **Figure 11**) based on streams generation capabilities of IBM InfoSphere and Spark to emulate data from a large number and a large variety of sensors. Such generated data will be sent into and processed by the applications residing on the IoT/M2M
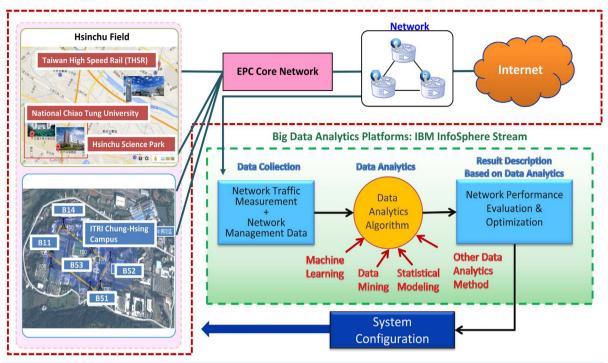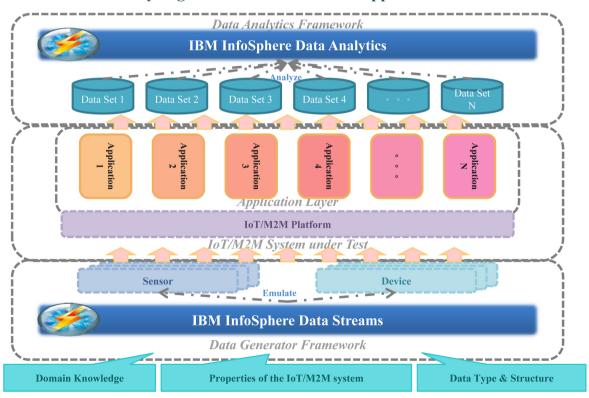
Figure 10. Big Data analytics for Traffic engineering of ITRINET.

Figure 11. IoT platform integrated with data generation and data analytics.

platform (illustrated in the middle part of **Figure 11**). Finally, the data sets produced by the IoT/M2M applications are analyzed by data analytics engines in IBM InfoSphere and Spark (illustrated in the upper part of **Figure 11**).

The effectiveness of our data generation methodology is first verified in a people management system of a large factory environment. With our methodology, we are able to test our factory management IoT/M2M application more efficiently without actually deploying a large number and a large variety of sensors. This application development uses both IoT and Big Data Analytics.

4.4. Application on Data Analytics for Traffic Generated by Testing & Measurement

The automatic Testing & Measurement (T&M) of DUT (Device under Test, including 4G/LTE SD, UE/CPE) for 4-Stage (namely, conformance, interoperability, operator-IOT and field trials) testing is shown in **Figure 12**.

We apply the data analytics learning techniques for the classification of the DUT types by series numbers or id. Once the type is identified, the parameters for feature extraction can be measured by the testing and measurement equipment. An example for DUT classification is shown in **Table 3**. In this example, the DUT type relational table is a 4×4 table with 4 fields: 1) DUT Device Type, 2) UE capability Information (FGI), 3) DNS (TTL), and 4) HTTP (User Agent). This device relational table is created by applying multivariate statistical analytics on the network traffic flow and LTE UE connection related information. Then by extracting specific features, we are able to identify the types of DUT such as Dongle, Smartphone, CPE, etc.
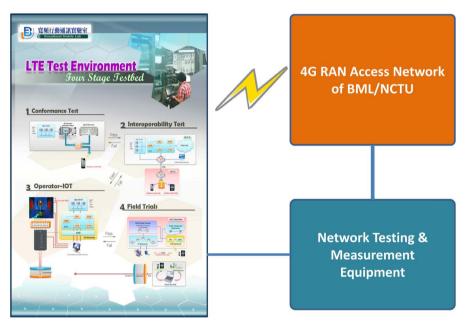
Figure 12. Data generation by endpoints from testing & measurement.

Table 3. DUT relational table.

DUT Device Type	UE FGI	DNS-TTL	HTTP-User Agent
Dongle A	0100 0110 0000 0101 0001 1000 0000 0000	128	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.134 Safari/537.36\r\n
Dongle B	0101 1111 0000 1111 1111 1100 1000 0000	63	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.134 Safari/537.36\r\n
Smart Phone	0101 1110 0000 1101 1101 1000 1000 0000	43	iPhone6,2/8.3 (12F70)\r\n
CPE	0101 1111 0000 1111 1111 1100 1000 0000	127	Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.134 Safari/537.36\r\n

Another example of applying data analytics in T&M traffic is to evaluate LTE QoS. When evaluating LTE QoS, it is required to setup testing scenarios with different QoS levels. Traffic Flow Template (TFT) (as shown in **Table 4**) is designed to filter packets into correspondent bearers, either a default bearer or dedicated bearer, in LTE network. Each bearer has its own QoS level. We can establish a dedicated bearer of guarantee bit rate (GBR) for applications such as VoIP, or a default bearer of basic QoS level for applications like file transfer.

What TFT does is to filter packets into correspondent bearers according to packets' IP address, port number, protocol, direction. The information, however, is insufficient to differentiate among web file download, FB news browsing, and Line chat. We are unable to give different QoS levels to these applications. To tackle this problem, we propose a new architecture which integrates deep packet inspection (DPI) with TFT to provide a higher granularity of QoS levels for applications. When some traffic flow has been identified as a certain application type, TFT will be informed to update rules of packet filters. With this improved method, the traffic flow can be delivered through the bearer of suitable QoS level.

4.5. Applications on Data Analytics for Traffic Flow Created by APP's of SD

To fulfill the Quality of Service (QoS) requirements from users, it is important to make effective use of the network resource. We can optimize the performance of a network by applying data analytics to traffic engineering. In particular, it is important to classify mobile applications traffic intended by the user with data analytics.

We propose a HMM-based (Hidden Markov Model) model to classify the mobile applications. By surveying related work, we have realized that there are different handshake patterns of well-known application protocols. Also, according to the observation of some specific mobile applications, we discover that every mobile Internet service has its unique negotiation process at the beginning when service starts. HMM was widely used to recognize time-dependent sequences and find out unknown patterns of data, such as speech recognition, handwriting recognition.

Table 4. Traffic Flow Template (TFT): Use to specify the packet filters associated.

UL Packet Filter ID	DRB	Packet Filter Evaluation Precedence	Protocol Number (IPv4)/ Next Header (IPv6)	Remote Address and Subnet Mask	Single Local Port (UE)	Local Port Range (UE)	Single Remote Port Range (NW)	Remote Port Range (NW)	IPSec SPI Range	Type of Service (IPv4)/ Traffic Class (IPv6) and Mask	Flow Label (IPv6)
1	DRB2	6	6 (TCP)	IPv4: 172.168.8.0 [255.255.255.0] IPv6: 2001:0ba0::[ffff:ffff::]	60051	-	-	-	-	-	-
2	DRB2	2	17 (UDP)	-	-	-	60201	-	-	-	-
3	DRB3	7	6 (TCP)	IPv4: 172.168.8.0 [255.255.255.0] IPv6: 2001:0ba0::[ffff:ffff::]	-	60100:60200	-	-	-	-	-
4	DRB3	3	17 (UDP)	-	-	-	-	60300: 60400	-	-	-
5	DRB3	5	50 IPSec (ESP)	-	-	-	-	-	0x0F80F0000	-	-
6	DRB3	1	-	-	-	-	-	-	-	00101000, Mask = 11111100	-
7	DRB3	6	-	-	-	-	-	-	-	-	-
8	DRB1 (default bearer)	255	6 (TCP)	-	-	-	-	-	-	-	-

In our method, we extract the packet size sequence and packet transmission direction sequence of the first 20 packets to train the HMM model. **Figure 13** shows our designed model structure of HMM. X_n is the hidden variable. It represents the transmission states which cannot be observed directly. Because of the unknown transmission states, we need to use observation variable O_n^1 and O_n^2 as training features to build the mobile application models. Here, "n" is the sequence number of packets in a traffic flow. O_n^1 is the observation symbol of packets transmission direction. O_n^2 is the observation of packet size that quantizes to a certain range of packet size. We quantize smallest packet size to number 1 and largest packet size to number 8. The rest of packet sizes are then divided into six groups.

Figure 14 shows the main process of our classification system. First, we process the collected packets by reading from original PCAP files, and extract the necessary field of packet header, including source IP, destination IP, source port number, destination port number, packet timestamp, and packet size. Second, we use packet
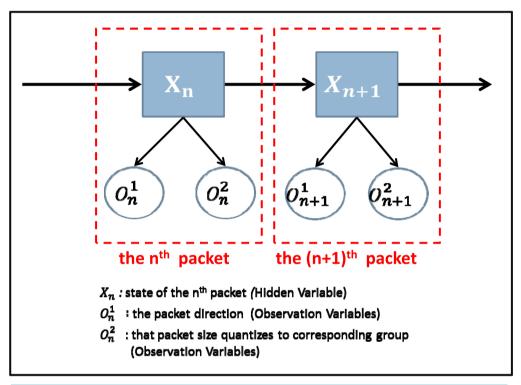
Figure 13. Classification model of Internet application traffic.

1. PCAP File Processing
Original packet → processing data
- SRC_IP / DST_IP
- SRC_PORT / DST_PORT
- Timestamp
- Packet size

2. Feature Extraction
First 20 packets of each flow
- Packet Direction
  1: Client → Server
  2: Server → Client
- Packet size
  Quantize to corresponding group

3. Learning Process
HMM Parameter Estimation
- Using EM algorithm
Application Training Model
- Each application has its own model

$$\lambda^1 = \{\pi_1, A_1, B_1\} \dots \lambda^N = \{\pi_N, A_N, B_N\}$$

4. Recognizing Process
Find Maximum Probability
- Observation of testing data: O^{test}

$$P(O^{test} \mid \lambda^1) \dots P(O^{test} \mid \lambda^N)$$

Max value of all training model

Figure 14. The process of our proposed classification system.

header information (5-tuple: source IP, destination IP, source port number, destination port number, and protocol) to process packets with the same 5-tuple information into a unit of mobile application traffic flow, and extract the features of each application flow. Third, in learning process, we use the extracted features which are quantized into corresponding symbol sequence to training the HMM-based application models for different applications. Finally, we can identify the new traffic flow by finding out the maximum value of log likelihoods derived from different application models.

5. Conclusion

This paper presents a forward looking view of the convergence of IoT, big data, cloud, SDN technologies along with the arrival of 5G mobile broadband networks. We intend to demonstrate the technical relationships of those technologies and the compelling programs and applications that can be created when two, three or more of those technologies converge. Due to the nature of fast evolution of ICT and the ongoing innovation of those five technologies, this paper should be updated on annual basis to keep the related information up to day with the ICT major trends.

Acknowledgements

References

[1] Lin, B-S.P. (2015) The Roles of SDN, Cloud, IoT, & Big Data, in 5G Networks and Their Technical Relationships & Applications. Invited Talk, IEEE VTS APWCS 2015, Singapore.

[2] Alkhatib, H., *et al*. (2015) IEEE CS 2022 Report. IEEE Computer Society, IEEE.

[3] Neira, E.M. (2015) IEEE ComSoc CTN Special Issue on Ten Trends That Tell Where Communication Technologies are Headed in 2015.

[4] Tucker, L. (2013) On the Cloud of the Future. Cloud Computing Summit, San Jose.

[5] Kleiner, T. (2015) 5G Research in Horizon 2020.

[6] Thadani, A. and Andreoli, P. (2014) Realizing Business Value from Convergence of IoT + Big Data Technologies. 2014 *Midwest Architecture Community Collaboration* (*MACC*) *Conference,* Minneapolis, 13 November 2014.

[7] Dinh, H.T., Lee, C., Niyato, D. and Wang, P. (2013) A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches. *Wireless Communications and Mobile Computing*, **13**, 1587-1611. http://dx.doi.org/10.1002/wcm.1203

[8] Lin, B-S.P., *et al.* (2013) The Design of Cloud-based 4G/LTE for Mobile Augmented Reality with Smart Mobile Device. *Proceedings of* 2013 *IEEE International Symposium on Mobile Cloud*, *Computing*, *and Service Engineering* (*IEEE Mobile Cloud* 2013), Redwood City, 25-28 March 2013, 561-566. http://dx.doi.org/10.1109/sose.2013.57

[9] Lin, B.-S.P., Tung, L-P., Hsieh, I-C., Liu, T-H. and Chou, S-Y. (2015) The Design of Big Data Analytics for Testing & Measurement and Traffic Flow on an Experimental 4G/LTE Network. *Proceedings of IEEE WOCC* 2015, Taipei, 23-24 October 2015, 40-44. http://dx.doi.org/10.1109/wocc.2015.7346173

[10] Lin, B-S.P., Tung, L-P., Hsieh, I-C. and Chou, S-Y. (2015) Performance Estimation of MAR for Outdoor Navigation Applications on an 5G Mobile Broadband by Mobile Smart Devices. *Proceedings of IEEE APWCS* 2015, Singapore, 19-20 August 2015, 1-6.

[11] Lin, F.J., Ren, Y. and Cerritos, E. (2013) A Feasibility Study on Developing IoT/M2M Applications over ETSI M2M Architecture. *Proceedings of the* 1*st International Workshop on Internet of Things Technologies* (*IoTT*), Seoul, 15-18 December 2013, 558-563. http://dx.doi.org/10.1109/icpads.2013.100

[12] Chen, H.W. and Lin, F.J. (2014) Converging MQTT Resources in ETSI Standards Based M2M Platform. *Proceedings of IEEE iThings*, Taipei, 1-3 September 2014, 292-295. http://dx.doi.org/10.1109/ithings.2014.52

[13] Cerritos, E. and Lin, F.J. (2014) M2M-Enabled Real-Time Trip Planner. *Proceedings of the* 2*nd International Workshop on Internet of Things Technologies* (*IoTT*), Hsinchu, 16-19 December 2014, 886-891. http://dx.doi.org/10.1109/padsw.2014.7097902

[14] Lusiarta Putera, C.A. and Lin, F.J. (2015) Incorporating OMA Lightweight M2M Protocol in IoT/M2M Standard Architecture. *Proceedings of IEEE World Forum on IoT*, Milan, 14-16 December 2015, 559-564.

http://dx.doi.org/10.1109/WF-IoT.2015.7389115

[15] Adrianto, D. and Lin, F.J. (2015) Analysis of Security Protocols and Corresponding Cipher Suites in ETSI M2M Standards. *Proceedings of IEEE World Forum on IoT*, Milan, 14-16 December 2015, 777-782. http://dx.doi.org/10.1109/wf-iot.2015.7389152

[16] Lin, F.J. and Chen, H. (2015) Improving Utilization and Customer Satisfaction of Parking Space with M2M Communications. *Proceedings of IEEE World Forum on IoT*, Milan, 14-16 December 2015, 465-470.

[17] Lin, F.J., Tsai, W., Cerritos, E., Lin, B.-T. and Hu, W.-H. (2015) Identification and Analysis of Charging Factors in M2M Communications. *Proceedings of IEEE World Forum on IoT*, Milan, 14-16 December 2015, 160-165. http://dx.doi.org/10.1109/wf-iot.2015.7389045

WILEY | Hindawi

Research Article

A Model of Socially Connected Web Objects for IoT Applications

Sajjad Ali [ID],[1] **Muhammad Golam Kibria** [ID],[1] **Muhammad Aslam Jarwar** [ID],[1]
Hoon Ki Lee,[2] **and Ilyoung Chong** [ID][1]

[1]*Department of Information and Communications Engineering, Hankuk University of Foreign Studies, Yongin-si, Republic of Korea*
[2]*Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea*

Correspondence should be addressed to Ilyoung Chong; iychong@hufs.ac.kr

The Internet of Things (IoT) is evolving with the connected objects at an unprecedented rate, bringing about enormous opportunities for the future IoT applications as well as challenges. One of the major challenges is to handle the complexity generated by the interconnection of billions of objects. However, Social Internet of Things (SIoT), emerging from the conglomeration of IoT and social networks, has realized an efficient way to facilitate the development of complex future IoT applications. Nevertheless, to fully utilize the benefits of SIoT, a platform that can provide efficient services using social relations among heterogeneous objects is highly required. The web objects enabled IoT environment promotes SIoT features by enabling virtualization using virtual objects and supporting the modularity with microservices. To realize SIoT services, this article proposes an architecture that provides a foundation for the development of lightweight microservices based on socially connected web objects. To efficiently discover web objects and reduce the complexity of service provisioning processes, a social relationship model is presented. To realize the interoperable service operations, a semantic ontology model has been developed. Finally, to evaluate the proposed design, a prototype has been implemented based on a use case scenario.

1. Introduction

The Internet of Things (IoT) envisions billions of objects connected to the Internet that continuously generate data about the physical environment. Several state-of-the-art applications can be built using the information and services provided by the pervasive and heterogeneous IoT objects. However, building applications based on these objects raises some challenges. Examples include how the growing number of objects will interact or coordinate to render valuable services, how the complexity generated by the coordination of objects can be handled, and how a large amount of diverse data sensed from the surroundings can be managed properly.

Numerous promising solutions have been provided to handle the complexity and the heterogeneity of IoT objects. One of the prominent approaches is to develop virtual environments to resolve the issues caused by the heterogeneous objects. Virtualization has become a key component in many IoT architectures, in the form of either virtual objects or

virtual entities. A virtual object (VO) is defined as a digital representation of a real-world object [1] and provides a way for the IoT services to discover and compose features that cannot be achieved directly with a real-world object (RWO). It has been realized that issues such as heterogeneity of dissimilar objects and scalability can be effectively resolved using VOs. Hence, a crucial role of a VO is to bridge the gap between the physical world and the virtual world by hiding the physical characteristics of an IoT device and acquiring and interpreting its data and context.

Another way to reduce the complexity caused by a large number of objects is to link these objects in a social network [2], which provides a more exact solution than using a sole individual object. This principle has given birth to the concept of Social Internet of Things (SIoT), which is emerging as a new paradigm with the merger of social networks and IoT [3–7]. It has been realized as an efficient way to facilitate the development of complex future IoT applications. It supports sharing of information generated by

the people and the devices based on the social relationships. These relationships further promote efficient discovery of the objects and effective service compositions.

Let us have a glimpse of the application scenarios in which SIoT will make a profound effect on our daily life in the near future. Imagine a group of vehicles that form a mobile social network on the move when heading towards the same location or when they are bound in a common relationship. Vehicles may form diverse social communities; for example, small cars create a social network to share available parking space information or bicycles share information about the vacant paths using a network [8]. Vehicles commuting from home to office can join a social network to share traffic congestion experience, accident warnings, or other common information (e.g., a meeting plan) with colleagues by accessing their vehicles' social network. Let us analyze another SIoT use case where a user enters a public museum with an IoT environment, which incorporates smart objects that exploit social relationships to share information. At the entrance to the premises, the user's smartphone is prompted with a beacon signal to recommend a service app installation. When the user grants the permission, an app is installed and a social agent acquires his profile and friendship details. The app incorporates a microservice which uses a smartphone object to establish a social relationship with the museum visit service. The museum is equipped with a network of smart objects that monitor the user's location and provide services to enhance his/her experience such as personalizing the displays, assisting in the navigation, and recommending the relevant services.

To support the above-mentioned use cases, most of the current IoT architectural approaches are not suitable as they are based on the traditional monolithic approaches, which further hinder the development of efficient, modular, and independent services that cannot scale well with the increasing user's service demands, as already witnessed in many studies [9–12]. Using monolithic approaches, we cannot fully achieve the benefits offered by SIoT including efficient information discovery, improved scalability, and simplified interconnection of objects.

These limitations motivate the current paper to define a design for the development of SIoT based services with efficient and lightweight mechanisms to exploit SIoT features for improved service provisioning. The proposed design is based on the microservices concept which promises a more solid practice of SOA. There is no particular definition of microservices architectural style; however, it is explained in [13] by Martin Fowler as "an approach to developing a single application as a suite of small services, each running in its own process and communicating with lightweight mechanisms, often an HTTP resource API." Following the microservices based proposed design, SIoT applications can be developed, deployed, and maintained more efficiently and independently, which will promote system modularity and interoperability.

Moreover, a SIoT design needs to facilitate the efficient information discovery based on the social relationships among objects. To achieve this, incorporating the existing social relationship models in the proposed design can be useful. However, applications with intelligent service features require dynamic object selection. Therefore, objects need to acquire the ability to infer new relationships for interconnection with other objects in the system. This view further motivates the present article to develop a social relationship model based on the semantic ontology. The utilized ontology-based approach is highly useful to dynamically infer new social relationships for efficient service provisioning.

Furthermore, to fully achieve the benefits of both, the SIoT and microservices, selecting an IoT environment is necessary where not only can the social interconnection among heterogeneous objects be utilized but also the lightweight services can be developed to apply these relationships for effective IoT service provisioning. To fulfill this goal, the Web of Objects (WoO) [14] platform has been selected. According to the ITU recommendation (ITU-T Y.4452) [15], "the Web of Objects (WoO) is a realization way of the IoT services, where virtualized objects (i.e., virtual objects (VOs) and composite virtual objects (CVOs)) are connected, controlled, and incorporated with resources to facilitate the development, deployment, and operation of IoT services on the World Wide Web."

To exploit the opportunities provided by the above discussed technological advancements, the main contributions are highlighted as follows:

(i) This article contributes a novel design based on the microservices concept with lightweight and modular services to support the development of SIoT applications. The proposed design involves mechanisms to enhance object discovery and to reduce the magnitude of complexity generated due to the huge number of objects.

(ii) A social relationship ontological model is developed, which helps identify the hidden and unidentified relationships among objects with reasoning mechanisms.

(iii) A use case with the implementation prototype based on the WoO platform is developed to exploit the social relationship model, supporting the proposed design for efficient IoT service provisioning.

Further, to support the above contributions, virtual objects, serving on real-world objects, have been developed, which can be reusable in other services as well. CVOs are implemented, where each update of the information in the real-world object is reflected in the VO and propagated to the CVO. Also, the Social IoT notion is incorporated into virtual objects to provide easy discovery and efficient execution while maintaining collaboration using social connections. Moreover, an experimental analysis has been carried out to evaluate the discovery time of VOs and the time required for service execution, with or without social relationships among virtual objects.

The remainder of this article is organized as follows. In Section 2, the related work is described. Section 3 presents the proposed design of the social web objects accompanied by the social interaction model. In Section 4, an IoT environment

use case is explained with the details of the prototype implementation and discussion of the performance evaluation. Finally, Section 5 provides a conclusion of this paper.

2. Related Work

This section presents the related work with respect to the SIoT and microservices. Further, the significance of both technologies pertaining to the IoT environment has been discussed.

2.1. Social Internet of Things (SIoT). The concept behind SIoT is to enable smart objects to create a social network similar to a human social network. The objective is to exploit the social relationships among objects in an IoT environment to facilitate the effective information discovery, to promote the scalability, to enhance the interactive communication between objects that are friends, and to achieve trustworthiness [3].

The initial notion of a social network of objects was proposed by [16], which discussed how the wireless devices form social connections on the temporary bases that are controlled by the node owners. As the work was carried out before the introduction of IoT, it did not consider the SIoT concept. Later studies [5] investigated the objects' social interaction where objects formed social networks and communicated with each other based on the rules already set.

The idea of using the Web of Things in a social ecosystem was presented by [17, 18], where things were shared using the social network infrastructures, for example, Twitter, Facebook, and LinkedIn. These implementations consisted of objects that communicated either through the built-in embedded servers or through smart gateways. Further, in these settings, the web-enabled things owned by a person provided an interface to his/her social group of friends so that they can interact with the things using a social network.

Lysis [19] is another good example of Social IoT platform based on cloud infrastructure. It uses the platform as a service (PaaS) model and focuses on the deployment of applications in the cloud space. However, dynamic inferencing of new social relationships from existing relationships is not supported by this system. Moreover, making a modular interoperable design by employing concepts like microservices and semantic web technologies is not considered in this work.

Paraimpu [20] provided a social Web of Things platform to connect virtual and physical things to the web. In this system, the social concept only relates to humans by providing them with the capability to share things with each other using a human social network. Also, virtual things are not considered as VOs but rather viewed as services on other IoT platforms.

Using SIoT at the edge of the network was investigated by [21]. In this system, an approach was proposed to cope with the communication delay due to the objects being remotely located in the cloud. This approach exploits the computing resources at the network edge to deploy virtual objects.

The ThingSpeak solution was used as the basis for the implementation of an early SIoT platform [22]. This platform provided the object social behavior functions incorporated at the centralized server. In this implementation, functions such as the creation and the management of the social relationships were developed. However, VO concept was limited only to the records in the remote database.

Another featured SIoT framework was contributed by [4]. This work presented how to combine services, devices, users, and their social interaction to enable interactivity, discovery, and recommendation of services. Social relationships have also been investigated in communities [23], where a proposed framework enables the identification of communities in the social networks.

Although most of the research contributions in the SIoT domain incorporate a notion of the social association of objects, however, limited information has been provided about the social relationship model and the details on how that can be used by IoT services. Furthermore, existing approaches are lacking lightweight and modular mechanisms which can be utilized by services to exploit social relationships efficiently. In our approach, we have developed a semantic ontology-based social relationship model that provides the capability to infer new connections in a network of social objects. Another distinction of the proposed system is a microservices based design that utilizes the developed social model to provide lightweight, modular, simple, and interoperable IoT services.

2.2. Microservices. The microservices architecture is emerging as a new trend among the practices of developing distributed web applications. To develop SOA based systems, microservices have become a prominent approach in real-world settings [24]. In the microservices based architectural pattern, each application incorporates a collection of small services which execute independently and use lightweight communication mechanisms [9]. The microservices are targeted for doing one thing well at a time based on the principle of single responsibility [25], as defined by Robert C. Martine: "Gather together those things that change for the same reason and separate those things that change for different reasons" [26]. Also, the microservices are conceived as autonomous entities, which means they can be changed and deployed independently of each other without requiring their consumers to change [24].

In the last few years, several IoT architectural designs were proposed and developed. Some of these designs provided innovative features such as the European FP7 project iCore [27] which proposed a cognitive framework for IoT application development. This project provided an architecture with the concept of virtual objects and their composites. However, in iCore, services are designed using a traditional SOA based monolithic style, whereas the microservices based architecture provides a better option to develop IoT services. Furthermore, in iCore, VOs lack the mechanisms to semantically represent real-world objects (RWOs), which limits extensibility and interoperability.

To support IoT applications, microservices based architectures are being proposed in many research initiatives and applied in several research projects. In [28], microservices are used for M2M applications, realizing the fact that monolithic approaches for M2M cannot provide a real solution. Another

work [10] provided an analysis of using microservices architecture for large-scale distributed applications. This work demonstrated that significant benefits can be achieved as compared to generic SOA approaches when using microservices in an IoT platform for smart city applications.

Moreover, Almanac FP7 EU Project [29] provides smart city services based on the IoT environment. The project uses microservices to employ scalability in a horizontal and vertical fashion. Also, in the industry, microservices have become a solution for developing large-scale applications. Netflix, Pivotal [30], and Amazon use microservices in their software bases.

The increased importance of microservices for the IoT applications is due to the fact that they simplify complex systems. By dividing a system into smaller parts, higher cohesion and lower coupling can be achieved, which makes it highly scalable. In scenarios where requirements keep changing continuously, microservices can help make a system easily modifiable. Moreover, the major benefits that are envisaged when using microservices in an IoT environment are as follows: Microservices leverage scalability to provide a highly decoupled pattern and can allow scaling individually. Their usefulness can be realized as when the demand for the requirement increases the system complexity also rises. In this case, a system can be supported by creating new instances of existing services. Microservices scalability [9] fits in three-dimensional space, that is, horizontal scalability (typical scalability), vertical scalability (splitting different individual microservices), and z-axis scalability (splitting similar things, such as DB partitioning). Microservices architecture supports a plug-and-play behavior where system components become loosely coupled. To acquire new functionality or replace a failed service, pull the plug from one microservice and plug into a new microservice [31]. One of the major problems that IoT is facing today is how to deal with the heterogeneity of incompatible solutions. To harmonize the heterogeneity, semantic web technologies with microservices are leveraged to provide interoperable exchange and communication of data. Microservices realize a decentralized and autonomous behavior, operating on their own priorities and schedule. This way, they provide several benefits for being utilized in an IoT design.

3. The Proposed Design of Social Web Objects

Recently, there have been several research efforts [3–6, 17, 20, 32] for defining the Social IoT. This new aspect minimizes the complexity generated by the communication of billions of real-world objects and enables sharing of information. The social notion can be incorporated into the WoO platform to provide a semantically rich base for IoT applications that can utilize real-world object relationships for efficient information discovery. Moreover, this section provides a brief introduction of WoO reference architecture in Section 3.1. A classification of social relationships among web objects (virtual objects and composite virtual objects) is discussed in Section 3.2 and a social relationship model is presented in Section 3.3. Further, the details of functional components for the proposed architecture are explained in Section 3.4.

3.1. Web of Objects (WoO) Reference Architecture. To enable the deployment of IoT services on the World Wide Web, WoO provides a reference framework. In other words, it realizes the IoT services in such a way that virtualized objects are interwoven with resources to support the development, deployment, and operations of IoT services [33, 34]. The WoO platform provides service functionality by merging VOs and features of web applications. Moreover, the WoO platform uses semantic web technologies to enable interoperability among heterogeneous resources. This realization provides the basis for the composition and harmonization of objects to provide smart services in an IoT domain. Building services based on WoO platform using diverse technologies, including microservices [35], social networks, and semantic web, helps reduce the complexity and fosters the easy and efficient development of IoT applications.

Virtualization has become a major concept in IoT to address the heterogeneity of diverse types of physical objects. VO is a digital representation of a real-world object that can be anything, living or nonliving, mobile or stationary, concrete or immaterial. In WoO, VO is defined as domain-specific semantic ontology based on the VO information model [36]. It is uniquely identified using URI and provides information updates on the representative real-world object. On the other hand, WoO also provides a notion of CVO which aggregates one or more VOs to enable service features that satisfy the application requirements. CVOs chain semantically interoperable VOs together. The WoO platform incorporates features to efficiently reuse existing VOs and handles the complexity generated by self-management and control mechanisms [37]. In the WoO, service level decides which CVOs and VOs will take part in creating an object mashup to fulfill service objectives. The service logic is saved in the form of templates and stored in the template repository [38–40]. The domain expert or knowledge engineer defines the service templates that are used to instantiate new services.

WoO reference architecture is shown in Figure 1. In this architecture, the service layer handles requests and provides several management functions, whereas the CVO level consists of functions to manage and instantiate CVOs or reuse the existing ones based on the provided service execution logic. New objects such as sensors are registered in the WoO platform using a registration function and their templates are generated to make them digitally available in the form of VOs. Another most important aspect is the semantic representation of data at three-layered WoO architecture [23, 39]. The database at each layer is supported with semantic web technologies.

3.2. Types of Social Relationships. Previous research [3] on SIoT has derived some basic relationship types among objects. These relationships are characterized as follows. *Parental Object relationship* (POR) exists between the objects that belong to the same batch, such as objects that are created in the same production process or at the same time by a common manufacturer. *Cowork object relationship* (CWOR) is the relationship among objects that are grouped together based on some commonly shared job to be done by them. Similarly, *colocation object relationship* (CLOR) is established
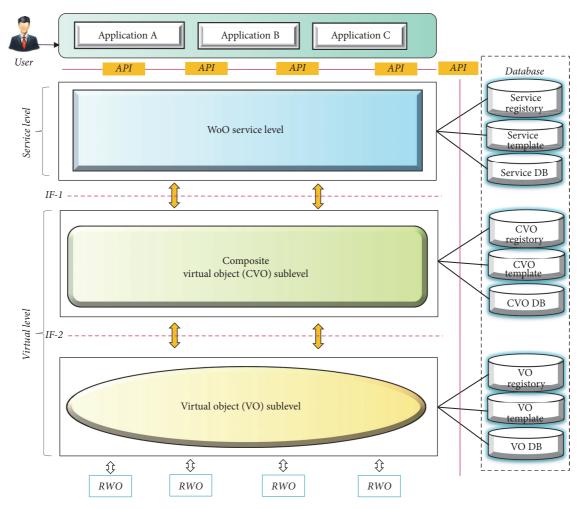
FIGURE 1: Web of Objects reference architecture.

between objects belonging to the same location such as home or office. Different or same types of objects can be combined based on the common location or premises. On the other hand, *Ownership Object Relationship* (OOR) is formed among objects belonging to the same owner. *Social Object Relationship* (SOR) is another kind of relationship that is created among objects when they come close to each other, either random in time or periodically. This relationship can be envisaged when the owners of objects come into contact with each other. Apart from the above relationships, in [41], the authors define the *Guardian Object Relationship* (GOR) in the Internet of Vehicles (IoV) scenario where on-board units of vehicles become children in relation to the super nodes of Road Side Units. This child and parent association gives a special meaning to a new hierarchical relationship.

In addition to the relationships defined in other studies, we have identified few more relationships suitable for some diverse scenarios. These include *Sibling Object Relationship* (SIBOR) that is created among objects that belong to a group of friends or family members. This relationship among objects extracts the property of trustworthiness based on the relationship of their owners. Another relationship type is the *guest object relationship* (GSTOR) which is formed between

objects that belong to the users in the guest role, for example, when a person visits a friend's home and gets the privilege as a guest. The same can be applied to the objects that move in guest relation (GSTOR) from one place to another; they will have some privilege of accessibility of information as compared to other objects. *Stranger Object Relation* (STGOR) applies to objects that encounter the presence of each other in an anonymous environment such as on the go or in the public environment. Similar to people who meet each other sometimes regularly but are anonymous to each other or are not fully aware of each other, this relationship can be used among objects to form different trust levels or to form strict restrictions to promote secure connections. Moreover, in *service object relationship* (SVOR), objects form a relationship while coordinating in the same service composition to fulfill a service request. Generally, in the proposed scheme, service request execution generates a mashup of objects, and objects that belong to the same mashup are assumed to be in SVOR.

We believe that the above defined new social relationships are beneficial in several distinctive scenarios. For example, SIBOR is more useful in scenarios where trust is highly important for establishing a connection. Unlike the more generalized relation POR, SIBOR is established

(i) Sibling Object Relationship (SIBOR)   (v) Parental Object Relationship (POR) [3]
(ii) Service Object Relationship (SVOR)   (vi) Colocation Object Relationship (CLOR) [3]
(iii) Stranger Object Relationship (STGOR)   (vii) Ownership Object Relationship (OOR) [3]
(iv) Guest Object Relationship (GSTOR)
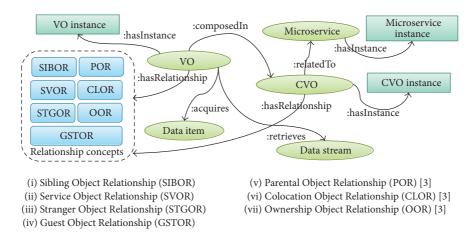
FIGURE 2: Ontological model representing web objects' relationships.

among objects from the same manufacturer having similar features but different behaviors. Similarly, GSTOR is suitable in scenarios where objects use relationships for privileged services. Imagine a person as a guest speaker in a conference with his smart devices, where he can connect to the network easily, and his personal devices are recognized as a guest (with GSTOR) in the venue network to avail free services like high-speed Internet, special notifications, and others. Stranger Object Relation (STGOR) is used in scenarios where objects are on the go, and to get some information, they have to compromise with a limited trust level. For instance, in a public transit system, one mobile node on the move wants to use crowdsourced information from another mobile node which is not fully trusted. Also, SVOR is useful in service composition specific scenarios, where object mashups are formed to facilitate service requirement, and it is also useful in a particular orchestration of objects.

To model information about microservices, social web objects, and their relationships, a semantic ontological model has been developed as depicted in Figure 2. Conceptualization of objects and their social relationships in semantic structures are highly beneficial to infer new connections among a network of objects with the help of reasoning techniques. The ontology facilitates the representation of objects, as well as what they measure in terms of observations, their processing, and functions. Web objects' relationship ontology contains the concepts that define each object in terms of a class. The data properties are used to represent object values and the object properties are defined to identify the link between two objects in the ontology. Major concepts in the ontology are microservices, CVOs, VOs, and their social relationships. Further, microservices have an individual instance that uniquely defines a microservice object and associated data with it. Similarly, CVOs and VOs are also instantiated and contain data properties to retain their values. With the increasing number of objects, a semantic ontology model is used to perform reasoning function that helps identify new connections. Further, to analyze the social relationship among objects (such as VOs and CVOs), their properties are provided in Table 1.

3.3. Social Relationship Model for Web Objects. At each level of service life cycle, objects form a social relationship with each other to accomplish a service task. Codifying the relationship among web objects can provide many benefits, such as efficient information discovery from related objects and better composition and reuse. To perceive social interaction between web objects, we assume every service is based on one or more microservices and other objects including CVOs and VOs. At VO level, VOs create several relationships with each other based on the RWO they represent. CVOs associated with microservices are incorporated in relationships as well.

Object-to-object associations are categorized as either vertical associations (interobject relations) or horizontal associations (intraobject relations) as shown in Figure 3. The relationships that flow from the bottom to the top in WoO based architecture are known as interobject relations (i.e., from VO to CVO or vice versa). On the other hand, relationships are known as intraobject relations if they are generated within objects at each level such as between one VO and another, or within one CVO to a relative CVO. The newly defined relationships among objects are maintained at different levels, which enables information discovery more efficiently. It is considered that not only will the relationships be maintained when two objects (i.e., VOs) are used in the same CVO or service, but also they will be maintained if objects are used in different CVOs or services.

As depicted in Figure 3, VO1 and VO2 bind in the SIBOR relationship, considering the fact that they belong to the same family of CVO1. Similarly, VO4, VO5, and VO6 are combined in a CWOR, as they are used in the same service. However, all of them do not have a common family relation that makes their trustworthiness restrictive to some level as compared to SIBOR. Moreover, though VO7 and VO8 are not used together in a service setup, still they form a STGOR relationship. This relationship also helps codify those VOs that have similarity in premises during a certain point in time regularly but they had never combined in any service scenario before. For example, if a person visits a subway regularly but stays very shortly, in such scene, the user's smartphone VO and subway station point

Table 1: Social relationship attributes.

Social object relationships types	Attributes/properties
Sibling object relation (SIBOR)	O_x *has relationship* with O_y → SIBOR, if objects' *ownership* is defined as: == SameFamily
Service object relationship (SVOR)	O_x *has relationship* with O_y → SVOR, if objects' *serviceStatus* is defined as: == SameServiceComposition
Stranger object relation (STGOR)	O_x *has relationship* with O_y → STGOR, if objects' *authentication* is defined as: == Anonymous
Guest object relationship (GSTOR)	O_x *has relationship* with O_y → GSTOR, if objects' *ownership* is defined as: !=SameFamily && == GuestFriend
Parental object relationship (POR)	O_x *has relationship* with O_y → POR, if objects' *creation* is defined as: == SameBatch ‖ == SameProduction
Colocation object relationship (CLOR)	O_x *has relationship* with O_y → CLOR, if objects' *proximity* is defined as: == Neighborhood ‖ *location* is defined as: == SameLocation
Ownership object relationship (OOR)	O_x *has relationship* with O_y → OOR, if objects' *proprietorship* is defined as: == SameOwner
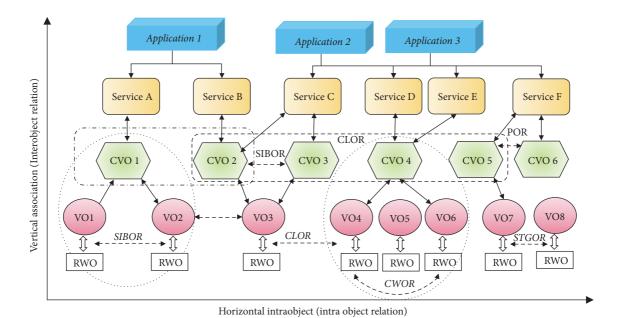
Figure 3: Hierarchical social relationships of web objects.

VO or subway security camera VO may establish STGOR. This may help reuse VOs in service scenarios where only for short duration a VO object is required. At CVO level, CVO2 forms two different relationships, which increases the level of connectivity of the CVO. CVO2 first shares a CLOR with CVO1 and CVO3, CVO4, and CVO5 together. Secondly, CVO2 is combined in a SIBOR with CVO3 and becomes in the same service family. Further, in the selected use case scenario described in Section 4, several other social interactions have also been elaborated.

3.4. Social Web Objects Architecture. SIoT envisages a system where the social framework will bring smart devices and people to interact with each other. By incorporating web technologies like SOA based microservices, IoT services can be rendered on top of the social framework. The proposed social web objects architecture (as shown in Figure 4) has been decomposed into three levels: service level, object

virtualization level, and aggregated object virtualization level. At the service level, to support social relationships within web objects, several microservices have been designed, which are discussed further in the following section. Moreover, the data management function in the proposed design helps each layer to interconnect with semantic databases. The SPARQL endpoints have been defined at the service and the object virtualization levels that expose interfaces to retrieve, store, and modify RDF graphs. Several interfaces allow the knowledge engineer, domain expert, and the developer to create service templates and VOs, as well as update the RWK model, user profile, and policies.

3.4.1. Service Level Functional Components. The service level functions handle operations from the request inception to the execution of services. This level is supported by some core functions common to be used in each service life cycle. These include the management function to handle service
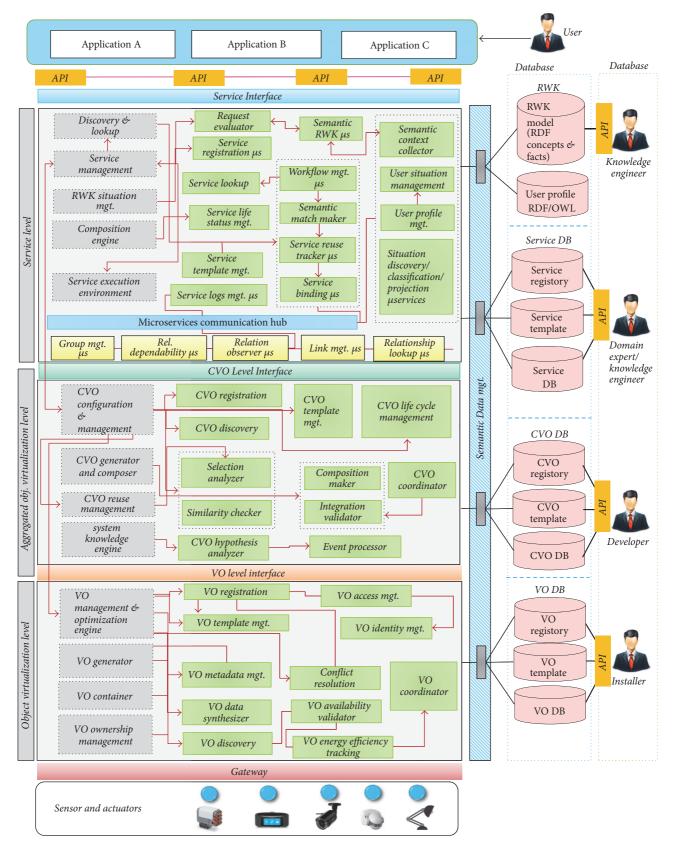
Figure 4: Social web objects functional architecture.

requests and initiate service provisioning processes, the service discovery function to select the available services, the composition engine to enable service mashups, and the Real-World Knowledge (RWK) and situation management function to acquire RWK and continuously maintain it. Besides these core functions, several microservices have been used to provide the plug-and-play feature in the proposed design; that is, in case of service failure, it is replaced with another one. These services include the Request Evaluator microservice which interprets a service request and matches a request query to the service template. The registration microservice registers newly added service objects to the registry. When a new template is instantiated, objects are put in service, and to know the existence of available service instances, they are recorded in the service registry. The lookup microservice is defined to search for the required service objects using the semantic representation. The representations available in RDF/OWL format are queried from the registry based on the request parameters. Service life cycle management keeps track of the states of service objects during their life cycle, whereas RWK management microservice is a semantic data processing service which processes facts about user preferences, profile, and situational information.

To realize service provisioning, it is always not possible to satisfy the service requirement with a single service function. Service composition and harmonization features are required to fulfill a demanding service request. Service composition is not a single operation; many microservices are used to compose the individual features. One of the major components of the composition is the Workflow Management (WFM) microservice which is responsible for decomposing the service requirement if no single service is available to satisfy the task. WFM microservice identifies the service objects using input and output interfaces. In this concern, it uses the matchmaking microservice to identify the approximate matches that can serve the required service functionality. Here, service objects are semantically represented in the form of ontologies. The semantic ontology alignment algorithm is used to match the service up to a certain threshold level that is defined by the matchmaking microservice. The reuse tracker microservice is used to identify the current service instance usage by the application; it incorporates a coordination mechanism before instantiating any new instance, and this helps in service reuse by many clients. Binding microservice enables tying of services selected by WFM; this function synchronizes input and output of services in a workflow to produce a single service output based on the generated composition plan.

Furthermore, user characterization is another important element to support the IoT services by matching user requirements more closely. The User Profile Manager microservice handles many facts related to the human user. These include user profile, preference, context, and policies. Semantic Context Collector microservice organizes the services requested by the user and the context in which these services were requested. This helps in identifying user interest with respect to the service context. User context or situation helps in service selection more efficiently if a user requests again with a similar service requirement. Some microservices have

been designed to handle user situation information; these microservices include situation discovery, classification, and recognition functions. Situation discovery is used to process sensor data that is provided by CVOs to prompt a situation, whereas machine learning algorithms are applied to remove the false values and data normalization is performed in case inappropriate triggers in the data exist. Situation classification incorporates the reasoning methods and rules to generate a relationship over detected events and refine the missing knowledge between the events. Situation projection analyzes the events data with the help of machine learning methods and provides an output in the form of predicted facts that are related to the situation being detected. Moreover, all the microservices coordinate using a communication hub to share information with each other.

A flow of service inception to service execution in the proposed architecture is depicted in Figure 5 where service management is responsible for initiating the above-mentioned functionalities and instantiating the microservices. This process further results in the generation of CVO/VO mashup graphs to be handled by CVO management (discussed in Section 3.4.2). Situational information has also been used to improve request analysis and service provisioning, whereas VO management (discussed in Section 3.4.3) acquires and aggregates data from sensors and other RWOs.

Social Relationship Management Microservices. To support the social relationships among web objects at each layer of the proposed system, relationship management microservices are defined, which support the codification of relationships among web objects (VOs and CVOs) and their management. These microservices are defined as follows. Group management microservice identifies the grouping of objects into specific sets based on the type and interaction of objects; it builds a social graph of the object relationships. Link management microservice incorporates mechanisms to maintain relationships among different objects and it includes several other subfunctions. The first subfunction is object selection which involves choosing candidate objects that are likely to form a particular association. The second subfunction is matchmaking service which provides exact relationship match based on the type of objects. The third subfunction is the association life cycle manager, which maintains the relationship status between objects; it checks whether they are in active relation or not. It keeps track of relationship validity and also the duration of relationships. Relationship observer microservice enables the object activities to be observed, and it also handles the mutual sharing, which defines what information an object is allowed to share or use of another object based on the relationship among them. Relationship dependability recognizer microservice identifies, maintains, and continuously improves the level of reliability among objects. If objects A and B are frequently interacting and always perform a task together that leads to successful execution of an application to serve the requirements, then they are more likely to form a reliable relationship with each other. The consistency of the relationship allows objects to share their information with fewer constraints as compared to unpredictable association among objects. Relationship
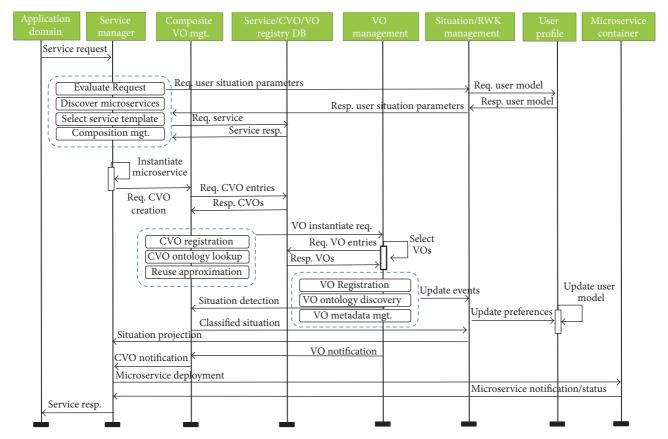
FIGURE 5: Request initiation and service execution flow in the proposed architecture.

lookup microservice incorporates the discovery algorithm (as given in listing of Algorithm 1) to discover the social relationships and maintains the list of all objects and their potential association rating. The rating defines the opportunity for objects to form a relationship with each other.

Moreover, the operation flow of social microservices has been illustrated in Figure 6, where VO management retrieves the social relationship among objects this is queried using link management, which uses a lookup service. Associations are maintained in a social relationship graph using semantic ontology and queried using the relation observer service. The social graph is also updated and maintained continuously with respect to changing links among objects, on the basis of which VOs are selected.

3.4.2. Functional Components in Aggregated Object Virtualization Level. Aggregated object virtualization (AOV) level involves all the necessary functions to instantiate, manage, and continuously monitor CVOs. Besides, it includes the interface required by the service level to interact with CVO components. Service level provides the mashup graphs of CVOs and VOs that are required to fulfill a service request. Aggregation level also supports some core functions; these are categorized as CVO configuration and management to decompose requests from service level and manage CVO instances, CVO generator to instantiate CVOs, composition function to form CVO and VO mashups, reuse management

to select existing CVOs that approximately match service requirement, and system knowledge engine to grow knowledge for the optimal use of CVO resources. Further, other aggregation level functionalities are as follows: registration function records the CVO entry into the semantic CVO registry in the form of RDF/JSON format. CVO includes ontologies that specify conditions to be applied on the VO data. Also, at this level, templates are created by a domain expert or knowledge engineer, which consist of functionalities associated with each CVO type. Template management module provides an interface to include templates to the CVO template repository and also supports the modification mechanism. CVOs are discovered using their semantic annotation via the discovery function. CVO selection analyzer function provides estimated selection of a CVO for reuse if there is no single CVO that can exactly provide the required service. This is supported by similarity checking mechanism to analyze the level of similarity before recommending for reuse. At aggregation level, another important function is event processing; event streams generated by object virtualization level are highly valuable to infer situational awareness and building RWK that is used by high-level functions. Event processing module facilitates processing of events and provides generated facts as output. Event processing is supported by CVO hypothesis analyzer that incorporates several hypotheses which are trained over data to build system level knowledge. Aggregation level also
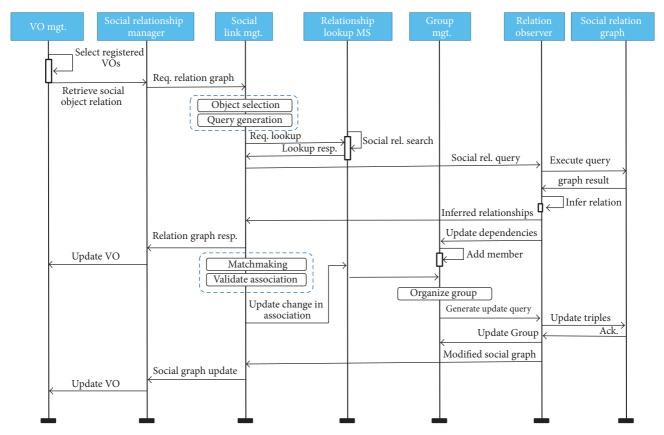
FIGURE 6: Sequence of operation required to establish a social relationship among objects.

provides composition function which provides the basis for merging multiple VOs to generate single service feature and validation function to validate the compositions of multiple VOs. Moreover, the coordinator helps in resolving conflicts in case of simultaneous access of single CVO by multiple services together.

3.4.3. Functional Components in Object Virtualization Level. Although IoT middleware virtualizes sensors and actuators in some form of virtual entities, however, in WoO, every living or nonliving thing can be represented in the virtual world. VOs are supported by an information model [14, 15] which includes the representation of RWO that can be ICT or non-ICT. In the proposed architecture, object virtualization level includes functions to create, maintain, and coordinate VOs such as VO generator, VO management, and optimization engine. VO registration function provides an interface to register VOs with the help of VO ownership management function where the VO ownership is maintained. Template management function handles templates or loads updated ones as provided by the domain expert. Metadata management maintains data associated with VOs; as VO is a digital representative of real-world entities, there is a high need to store metadata information to recognize the class of sensed data. Use of metadata provides the real value of the data in the form of the context stored within it. Furthermore, VO data synthesizer service handles the

discrepancies if found in the data corresponding to VO; this includes data manipulation and missing value rectifications. Conflict resolution function helps in case the same VO is used by multiple service instances, to resolve the conflicts. Also, VO access management function controls access for VOs in case access right information has been defined by the VO owner, whereas identification management module maintains identification information about VOs so that each VO can be uniquely identified in the system.

3.4.4. Social Relationship Discovery and Composition. To identify associations among web objects within a social relationship graph, an algorithm is elaborated briefly. In the following algorithm, the microservices that are in execution are retrieved first, represented as $\Sigma M\mu$; next, all the CVOs associated with each of those microservices are recovered in ΣMc and, further, the corresponding VO model ΣMv is iterated. In the next step, executing the query statement routine returns the results denoted as $\overline{\top}$. The process of finding the relationships is carried out through lines (6)–(19) in multiple nested iterations as shown in the pseudocode of Algorithm 1 listing.

Algorithm 2 facilitates the composition of objects based on social association types. This algorithm takes as input the set of object social relationships denoted by $\mathbb{Z}$ and registry entries of the objects that are currently available in the system, represented as $\mathbb{R}e$. After the object model denoted

Require: $(\Sigma\phi)$, $(\mathbb{R}_e)$
(1) **Output: $\hat{\mathbf{S}}$** (social relationship graph of objects)
(2) $\Sigma M_\mu \leftarrow$ Load Microservices in model
(3) $\Sigma M_c \leftarrow$ Load CVOs in model
(4) $\Sigma M_v \leftarrow$ Load VOs in model
(5) $\overline{\mathsf{T}} \leftarrow$ *executeQueryStatement* (M_μ, M_c, M_v, Q)
(6) **for All** μ in $\mu\overline{\mathsf{T}}$ **do** (iterate Microservice instance being used)
(7) 　　**for All** $O \in \mu$ **do** (check CVO and VOs used by the Services)
(8) 　　　　**if** $O_i ==$ output of C_R in a relationship set R_c **then**
(9) 　　　　　　add C_R to $\mathbb{I}_L$
(10) 　　　**else**
(11) 　　　　**if** $O_i ==$ output of V_R in relationship set R_v **then**
(12) 　　　　add V_R to $\mathbb{I}_L$
(13) 　　　　$\mathbb{N}_i \leftarrow$ tag (Assign relationship tag to each object entry)
(14) 　　　**else**
(15) 　　　　add O_i to $\mathbb{I}_U$
(16) 　　　**end if**
(17) 　　**end if**
(18) 　**end for**
(19) **end for**

ALGORITHM 1: Social relationship discovery.

Require: $(\Sigma\phi)$, $(\mathbb{R}_e)$, $(\mathbb{Z})$
(1) **Output: $\mathbb{C}$ (Composite Service)**
(2) $\Sigma M \leftarrow$ List all objects available (Microservices, CVOs, VOs) in model
(3) $\overline{\mathsf{T}} \leftarrow$ *executeQueryStatement* $(\Sigma M, Q)$
(4) $\mathbb{Z} \leftarrow$ *executeQueryStatement* $(\Sigma M, Q')$
(5) **for All** $\sigma i \in \overline{\mathsf{T}}$ **do**
(6) 　**for All** $\lambda i \in \sigma$ **do**
(7) 　　**if** $\lambda i ==$ *any* z in $\mathbb{Z}$ **then**
(8) 　　　add λi to $\mathbb{I}_{Lm}$
(9) 　　　store $\mathbb{I}_{Lm}$
(10) 　　**end if**
(11) **end for**
(12) **for All** $\varpi i \in \mathbb{I}_{Lm}$ **do**
(13) 　**if** $\varpi i ==$ *any* ϖi in $\mathbb{I}_{Lm}$ **then**
(14) 　　$\partial \leftarrow$ ranking (Assign rank to each service)
(15) 　　add ϖi **to** $\mathbb{I}_{LR}$
(16) 　　store $\mathbb{I}_{LR}$
(17) 　**end if**
(18) **end for**
(19) **for All** $\sigma j \in \mathbb{I}_{LR}$ **do**
(20) 　**if** $\sigma j(R\partial) > \kappa$ **then**
(21) 　　add σj to Wf
(22) 　　Pq $\leftarrow$ assign to queue
(23) 　　Sort Pq
(24) 　　Formulate Wf
(25) 　**end if**
(26) **end for**

ALGORITHM 2: Composition of objects based on social relationship.

as ΣM is loaded, the first step is to compose objects based on their relationship types. This way, the SPARQL queries are executed; the first query is denoted as Q to retrieve the available service templates and the second query is represented as Q' to extract the relationships associated with the services. From line (5) to (11), first, the social relationship is searched among objects and then the found relative objects are composed in groups. The design philosophy employed over here is to assign a ranking to each selected group based on the relationship type that helps in a more efficient
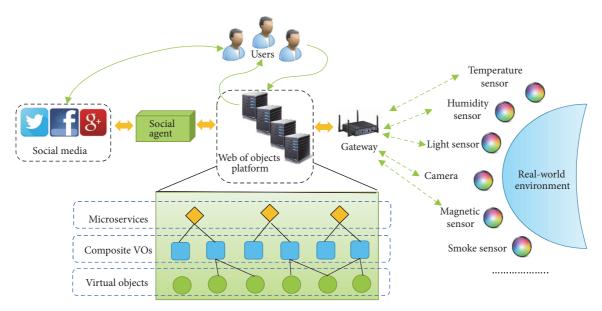
FIGURE 7: Use case scenario.

composition of service objects (from line (12) to (18)). The last part of Algorithm 2 will continue to iterate to compose the object composition workflow by checking first the assigned relationship ranking with the defined threshold, and then the selected composition workflow graphs are sorted and stored in queues for service execution.

4. Use Case and Prototype Implementation

In this section, first, to realize IoT service provisioning based on social web objects, a use case about IoT enabled museum environment has been presented and then the details on the prototype implementation are furnished.

4.1. Use Case Scenario. This use case is based on a user experience in a public museum where the IoT environment is already set up with the WoO platform incorporating web objects that use social relationship model and a social agent which obtains the user's profile and his friendship details from social media (as shown in Figure 7). The user enters the premises of the museum and his smartphone is prompted with a beacon signal to recommend service app installation where he accepts the request. The phone app incorporates a microservice which establishes a social relationship (GSTOR) with museum visit service objects. The system monitors the interaction of the user in the museum environment; the installed camera network and motion detectors detect the user presence and this feed is collected by the museum visit service. The user shares his current location with the system, which is used by the location navigator service to assist him in moving around the museum by providing the direction to the locations where different museum items are situated. It is worth noticing that the VOs for museum visit service and location navigator service are bound in CLOR relationship due to the same location points. The user shares his social network profile which is used to acquire the list

of friends and choices for things the user is interested in. This information is acquired by the recommendation service which suggests the best things to do with respect to the user's interests, making his visit most enjoyable. Recommendation service suggests facilities available in the museum such as places a user should visit, cafeteria food menu, and prices for the day. This service also shares with the user the past experiences of his friends at the museum to help him get most out of his visit. Moreover, personalization service provides special features such as adjusting contents on displays or optimizing the luminosity, to support users with disabilities. To enable these service features, objects exploit the SVOR relationship. Also, the fire management service maintains the status of all VOs that can detect an unusual fire situation in the museum. The temperature VOs acquire readings from temperature sensors and smoke detector VOs get a feed from smoke detection sensors. These VOs form a CLOR relationship with each other in the common proximity. Fire situation monitoring CVO acquires VOs data. One or more CVOs are managed by the fire management service. In case of a fire breakout, the fire management service executes with the data from the fire detection CVOs and the temperature and smoke detector VOs. Meanwhile, services incorporate high-level functionality to react on the data provided by virtual entities in the system. It is important to notice that sharing of information based on social relationships has become easy in the proposed use case. That is because the user information and device information are communicated with social links of web objects and are utilized by the services.

4.2. Proof-of-Concept Details. To evaluate the proposed design, a proof of concept has been furnished, which realizes services based on web objects as considered in the above use case. These objects include, for instance, fire situation monitoring CVO to handle emergency fire breakout situations, location navigator CVO to find the shortest path to the user
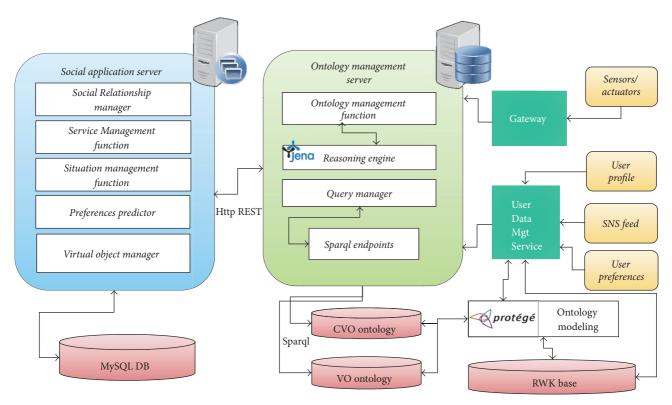
FIGURE 8: Prototype implementation components.

intended destinations, and user profile reader CVO which acquires user personal information from the social media profile. Several other CVOs have been developed that acquire diverse data and metadata from VOs where each represents a different RWO. These VOs include a user VO, smartphone VO, camera VO, wristband VO, motion detector VO, and others that are associated with the chosen environment (such as CO_2 sensor VO, humidity VO, temperature VO, and light VO). Each VO has several instances; for example, temperature sensor instances are represented as temperature-Sesor0001 and temperatureSesor0002. Further VO data is collected and analyzed by CVOs with the rules incorporated in the ontologies. CVOs composed in microservices help analyze different situations and initiate inference over data using inference engines. Considering the system requirement, several aspects have been evaluated, including identification of web objects for effective IoT service provisioning, finding the social relationships among web objects, efficient discovery of web objects based on these relationships, and composition and harmonization of web objects.

4.3. Prototype Implementation Details. To analyze the feasibility of the proposed architecture, a prototype on the discussed use case has been implemented. Figure 8 illustrates the components of the prototype, which constitute social application server (SAS), ontology management server (OMS), and databases to hold VO and CVO graphs that are supported with SPARQL endpoints. The gateway connecting sensors and actuators has been incorporated in the implementation. Additionally, user data management services are

implemented to get user profile and preferences history along with social network data.

The application server comprises five major components: social relationship manager (SRM), service management function (SMF), situation/context management function (CMF), preference predictor, and virtual object manager (VOM). The SMF is an entry point for the service request; it interprets and evaluates the service request. It is also responsible for matching the service request with the available template. SMF communicates with VOM to check the VO or CVO availability. A compiled list of VOs is provided to SMF which generates object mashups to satisfy the service request. The execution of VO mashup graph is done using microservices. The second function, VOM, maintains instances of CVOs and associated VOs and coordinates with OMS to use available ontologies in the system. It records the association of services and VOs and enables the reuse of VOs by more than one service; it also resolves conflicts using conflict resolution mechanism. VOM is implemented using RESTful web services. Further, SRM is implemented to maintain object relationships in the form of RDF concepts represented in ontologies. It also involves several microservices, each implemented for a specific task. For example, group management microservice is developed to group objects that have similar characteristics. Similarly, other tasks are distributed in different microservices; for example, link management microservice is used to identify the dependency among objects. Relationship observer microservice is implemented to monitor the objects when they change their state at runtime, whereas to search a web
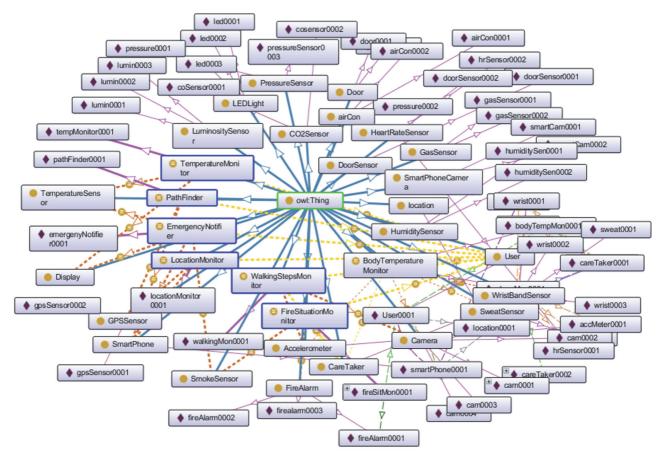
FIGURE 9: Ontological model of social web objects.

object relationship, the lookup microservice is used. The CMF function has been developed to process the situational information and it selects the most suitable VOs based on the context. The situational information contains the current location of the user, location of the RWOs, and the area under observation. CMF helps form a VO configuration specific to the user current situation. Additionally, to incorporate the real-world facts about the user and the situation, RWK base has been developed which is updated continuously. This also helps in adapting services according to the current situation. VOs provided data is accumulated in RWK model and the reasoning engine is used to infer new facts based on the existing data. The last function of social application server is the preference predictor which provides predictions by using machine learning methods. To implement the learning function, Weka Library [42] has been used. For the objective of demonstration, user preference patterns are collected from the social data and user preference history is used to predict the future patterns.

On the other hand, OMS deals with the semantic representation of VOs. The first functional component of OMS is the ontology management service, which is used to handle VO and CVO ontologies and provides an interface to use VO graphs. Further, to manage ontologies in the developed system, we have used Apache Jena framework [43], and to model

VO and CVO ontologies, Protégé [44] has been used. Models generated in Protégé are utilized by ontology management service and stored in OWL format in semantic data stores. To take full advantage of ontology-based implementation, a reasoning engine has also been used to get extra facts related to the modeled concepts. Reasoning engine helps to grow RWK facts stored in VO and CVO ontologies. For the proof of concept, the prototype has used Apache Jena inference engine along with Jena provided reasoner. For the deployment of persistent storage of the RDF data, Jena TDB has been used in our prototype, which is the Jena native high-performance triple store with API support [43]. Query manager is another important component of OMS which provides an interface to query CVO and VO ontologies saved in RDF/XML format, using SPARQL. Also, web services are implemented for the execution of SPARQL queries such as extracting and updating RDF triples in the database. Other than SPARQL query processing, semantic matching of existing concepts is also performed.

Moreover, to achieve interoperable relationships among web objects, a semantic ontology model has been developed as shown in Figure 9. The model constitutes the VO and CVO ontologies and their relationships. Several CVOs have been instantiated, for example, the location navigator, fire situation monitor, emergency notification, and many others that are

```
<!-- http://www.webofobjects.com/hufs/smartservice#01 -->
 <owl:Class rdf:about="http://www.webofobjects.com/hufs/smartservice#01#SmartMuseumService">
    <!-- http://www.webofobjects.com/hufs/smartservice#01#fireSitMon0001 -->
    <owl:NamedIndividual rdf:about="http://www.webofobjects.com/hufs/smartservice#01#fireSitMon0001">
        <rdf:type rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#FireSituationMonitor"/>
        <smartservice:CapturedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#cam0001"/>
        <smartservice:NotifyTo rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#fireAlarm0001"/>
        <smartservice:NotifyTo rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#smokeSensor0001"/>
        <smartservice:checkedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#User0001"/>
    </owl:NamedIndividual>

    <!-- http://www.webofobjects.com/hufs/smartservice#01#PathFinder0001 -->
    <owl:NamedIndividual rdf:about="http://www.webofobjects.com/hufs/smartservice#01#PathFinder0001">
        <rdf:type rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#PathFinder"/>
        <smartservice:checkedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#User0001"/>
        <smartservice:checkedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#surveillanceSystem0001"/>
        <smartservice:notifiedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#location0001"/>
        <smartservice:notifiedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#smartPhone0001"/>
        <smartservice:notifiedBy rdf:resource="http://www.webofobjects.com/hufs/smartservice#01#motionDetector0002"/>
    </owl:NamedIndividual>
 </owl:Class>
```

FIGURE 10: CVOs and VOs used in the ontology.

```
<!-- http://www.webofobjects.com/hufs/smart-service-ontology#smartPhone0001 -->

<rdf:Description rdf:about="http://www.webofobjects.com/hufs/smart-service-ontology#smartPhone0001">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
    <rdf:type rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#SmartPhone"/>
    <hasLocation rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#location0001"/>
    <hasOwner rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#User0001"/>
    <batteryStatus rdf:datatype="http://www.w3.org/2001/XMLSchema#string">70%</batteryStatus>
    <deviceName rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Samsung-S5</deviceName>
    <modelNumber rdf:datatype="http://www.w3.org/2001/XMLSchema#string">SHV-E77OS</modelNumber>
    <networkStatus rdf:datatype="http://www.w3.org/2001/XMLSchema#string">WIFI</networkStatus>
    <osVersion rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Andriod version 5.0.1</osVersion>
    <smartPhoneId rdf:datatype="http://www.w3.org/2001/XMLSchema#string">S5-00:33:55:11:00</smartPhoneId>
    <smartPhoneManufacturer rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Samsung</smartPhoneManufacturer>
</rdf:Description>
```

FIGURE 11: Smartphone VO used in the ontology.

based on the composition of numerous VOs. This development has been supported by OMS and semantic web tools such as Jena. Using OMS, CRUD operations are performed like create, update, and delete on CVOs and VOs using SPARQL. In the developed semantic relationships, two CVO instances, namely, "FireSitMon0001" and "PathFinder0001," have been used in the code snippet provided in Figure 10, for fire situation monitoring and location navigation, respectively. We can see the data properties and object properties associated with each object to maintain the relationship with other objects in the system. Further, to understand the VO representation, a smartphone VO "SmartPhone0001" is shown in Figure 11 with the defined properties. A list of VOs used in our prototype has been given in OWL description as shown in Figure 12.

In the test bed environment, the sensor to the gateway connectivity is enabled using ZigBee and BLE. Another source of information that has been incorporated to provide user-centric service capability is SNS feed, which consists of Twitter and Facebook data collected using social media APIs. On the other hand, user profile and preferences are updated using data management service, and this information is used to grow RWK base.

To test the prototype, an android application has been developed. After the user logs in to the app for WoO based assisted services, the app starts the configuration of services according to the user profile. A list of services is presented on the user's screen from which the user can select any service. These services include museum visit service, location navigator service, recommendation service, and emergency notification. The selected service starts collecting information with the interaction of the user or when the devices start pushing values about any change in the observed environment. Based on these observed parameters, alerts are generated by the app; for example, the museum visit service generates alerts about the available facilities in the museum and the environmental conditions including the indoor and outdoor temperatures. On the other hand, the location navigator service helps the user to navigate around the museum by showing directions to several visitor attractions. Moreover, the emergency notification service triggers alerts when a temporary fire is created to mimic an unusual situation such as fire breakout. The resulting fire increases the temperature reading which is detected by the temperature sensor. This change in reading is forwarded through the gateway and analyzed by the fire situation monitor CVO which checks its threshold value and generates an emergency notification on the user's app screen. Also, the recommendation service is used to generate suggestions about the museum events and the places to visit.

```
<!-- http://www.webofobjects.com/hufs/smart-service-ontology#deviceName -->

<rdf:Description rdf:about="http://www.webofobjects.com/hufs/smart-service-ontology#deviceName">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#DatatypeProperty"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#CO2Sensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#Camera"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#DoorSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#FireAlarm"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#GPSSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#GasSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#HeartRateSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#HumiditySensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#LEDLight"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#LuminositySensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#PressureSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#SmartPhone"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#SmartPhoneCamera"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#SmokeSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#SweatSensor"/>
    <rdfs:domain rdf:resource="http://www.webofobjects.com/hufs/smart-service-ontology#WristBandSensor"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</rdf:Description>
```

FIGURE 12: List of VOs used in the ontology.
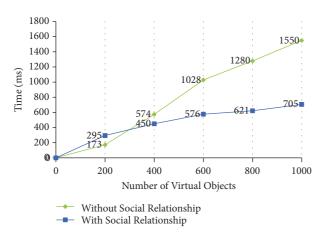
FIGURE 13: VO discovery time with the increasing number of virtual objects.
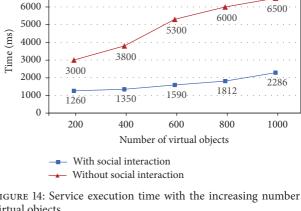
FIGURE 14: Service execution time with the increasing number of virtual objects.

4.4. Performance Evaluation. The proposed architecture has been analyzed from two different perspectives: system scalability and resource consumption. The results with respect to the experiments and the performance criterion are discussed in this section. The first experiment involves the analysis of the time required to discover the virtual objects in the proposed system with or without social relationship criterion as shown in Figure 13. As it is apparent from the results, initially, to form a social graph, VO discovery process incurs delay due to the communication for establishing social relationship links. However, direct lookup for VOs in the registry at the start of the discovery process comparatively requires less time, but as the number of VOs increases, the delay rises. Traversing the social links to discover objects as the number of virtual objects grows considerably reduces the total lookups in the registry database.

The second experiment provides an analysis of the execution time required for the CVOs as shown in Figure 14. To test this in a real environment, an android application

has been developed. The app allows access to the parameters from the available VOs. When the app requires retrieving observations, it iterates through VOs in the repository using a query interface. This requires high CPU utilization with more execution time required to retrieve the sensor values against each app generated request. This is due to the fact that the decision for the selection of VOs has been done at runtime; it requires more time to select the VO graphs, execute them, and get the relevant data from the VOs. However, when object interaction is based on the social relationship, it requires less time for the retrieval as friends in the social network share the data on other friends' requests, which results in less computation as compared with the previous scheme.

Moreover, Figure 15 depicts the time to discover VOs in four different types of services. These services are the museum visit service, the location navigator service, the emergency notification service, and the recommendation service. The discovery time varies from one service to another depending on the number of VOs used. However, as compared with the impact of social connections, it can be
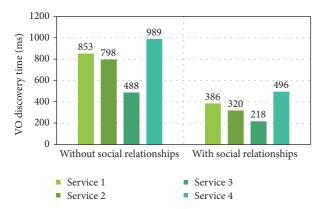
FIGURE 15: VO discovery time for four different services with or without social relationship utilization.
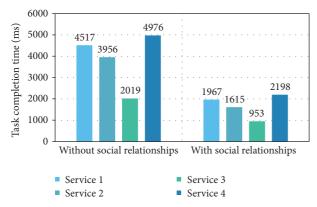
FIGURE 16: Task completion time required for each service with or without social relationship exploitation.

realized that the services can considerably reduce the time of object discovery and execution by using the relationship model. This is due to the fact that with social association among web objects a service does not have to query the registry for all VOs and rather VOs in the same social graph are utilized. Also, VOs collected observations are shared to save the service time for probing each VO. To view this, Figure 16 shows the overall time required for services to execute their tasks. An important factor to notice here is that the overall task completion time is also affected when social links are used to collect observations from VOs that are used in the service execution.

5. Conclusion

The Internet of Things is bringing the next technological revolution by connecting billions of objects on the earth and providing intelligent IoT services. However, it also carries two most important challenges: the first one is the complexity of handling a huge number of heterogeneous objects and the second one is how to deal with the monolithic approaches for providing services over existing IoT infrastructure. To address these challenges, we have proposed an integrated design based on the principles of SIoT, where the complexity

has been handled using socially connected web objects; to overcome the monolithic approaches, microservices that can compose new service features independently have been adopted. In the proposed design, a social relationship model has been presented, which enables the efficient discovery of web objects and reduces the complexity of service provisioning process with the algorithms to discover and compose web objects. A semantic ontology model has also been developed to realize the interoperable social interaction among heterogeneous objects. Finally, a prototype implementation based on a use case scenario has been demonstrated; to evaluate the system with respect to the performance issues, experimental results pertaining to the VO discovery and service execution time have been rendered.

Details of Notations Used in Algorithms

$\Sigma\phi$:   Service object ontologies based on semantic representation where ϕ is a replacement for M_μ which represents microservices M_c which represents CVO and M_v are VO ontologies

$\mathbb{I}_L$:   Collection or list of all relationships that are retrieved from service objects

$\mathbb{I}_U$:   List of the objects that do not have a relationship or are not associated

$\hat{S}$:   Social relationship graph of objects

U: Correlation union of templates

$\underline{\mu\overline{T}}$:   All service objects returned in response to the query request

$\overline{T}$:   All objects returned in response to the query request

μ: The single instance iterated from the collection of microservices

M: Data model based on the specified ontology, where ΣM_s represents service data model, ΣM_c represents context data model, and user profile model is represented as ΣM_u

Q: SPARQL query to retrieve the available service templates for user rating

Q':   SPARQL query to retrieve the relationship associated with services

$\mathbb{R}_e$:   Registry entries of service objects

C_R: Relationship associated with CVO

V_R: Relationship associated with VO

R_c: All possible relationships that can be associated with CVO

R_v: All possible relationships that can be associated with VO

O_i: Iterator object instance for iterating the list of CVO and VO objects

$\mathbb{Z}$:   Set of objects' social relationships

σi: Iteration item of the list of service objects

λi: An instance of iteration items in the list of service objects

$\mathbb{I}_{Lm}$:   List of all matched service items

ϖi: An instance of iteration items in the list of matched service items

$\mathbb{II}_{LR}$: List of all matched ranking
∂:   Ranking value assigned to a service object
σj: Iterator item for the ranking list
κ: Threshold to rank a service object
Wf: Workflow for the composition of service objects
Pq: Priority queue to store ranked object instances
$R\partial$:   Assigned relationship ranking.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
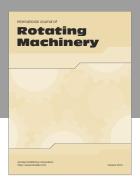
Acknowledgments

References

[1] M. Nitti, V. Pilloni, G. Colistra, and L. Atzori, "The virtual object as a major element of the internet of things: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1228–1240, 2016.

[2] J. Surowiecki, *The Wisdom of Crowds*, Anchor Books, 2005.

[3] L. Atzori, A. Iera, G. Morabito, and M. Nitti, "The social internet of things (SIoT)—when social networks meet the internet of things: concept, architecture and network characterization," *Computer Networks*, vol. 56, no. 16, pp. 3594–3608, 2012.

[4] A. M. Ortiz, D. Hussein, S. Park, S. N. Han, and N. Crespi, "The cluster between internet of things and social networks: Review and research challenges," *IEEE Internet of Things Journal*, vol. 1, no. 3, pp. 206–215, 2014.

[5] L. Atzori, D. Carboni, and A. Iera, "Smart things in the social loop: Paradigms, technologies, and potentials," *Ad Hoc Networks*, vol. 18, pp. 121–132, 2014.

[6] L. Atzori, I. Antonio, and G. Morabito, "Making things socialize in the Internet -Does it help our lives?" in *Kaleidoscope 2011: The Fully Networked Human?-Innovations for Future Networks and Services (K-2011), Proceedings of ITU. IEEE*, 2011.

[7] L. Atzori, A. Iera, and G. Morabito, "SIoT: giving a social structure to the internet of things," *IEEE Communications Letters*, vol. 15, no. 11, pp. 1193–1195, 2011.

[8] A. M. Vegni and V. Loscrí, "A survey on vehicular social networks," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, article no. A3, pp. 2397–2419, 2015.

[9] L. Krause, *Microservices: Patterns and Applications: Designing Fine-Grained Services by Applying Patterns*, 2015.

[10] A. Krylovskiy, M. Jahn, and E. Patti, "Designing a smart city internet of things platform with microservice architecture," in *Proceedings of the 3rd International Conference on Future Internet of Things and Cloud (FiCloud '15)*, pp. 25–30, Rome, Italy, August 2015.

[11] B. Butzin, F. Golatowski, and D. Timmermann, "Microservices approach for the internet of things," in *Proceedings of the 21st IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2016*, Germany, September 2016.

[12] N. Dragoni, I. Lanese, S. T. Larsen, M. Mazzara, R. Mustafin, and L. Safina, "Microservices: How To Make Your Application Scale," *Computing Research Repository*, 2017.

[13] "Microservices Guide Pages - Martin Fowler," 2017, https://martinfowler.com/articles/microservices.html.

[14] "Web of Objects. ITEA 3 · Project · 10028 , D2.1: State-of-the-Art Relevant to the Web of Objects," 2017, https://itea3.org/project/web-of-objects.html.

[15] "Y.4452: Functional framework of web of objects," 2017, http://www.itu.int/rec/T-REC-Y.4452-201609-P.

[16] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl5, and H.-W. Gellersen, "Smart-its friends: a technique for users to easily establish connections between smart artefacts," in *Ubicomp 2001: Ubiquitous Computing*, pp. 116–122, Springer, Berlin, Germany, 2001.

[17] D. Guinard, M. Fischer, and V. Trifa, "Sharing using social networks in a composable Web of Things," in *Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2010*, pp. 702–707, Germany, April 2010.

[18] M. Kranz, L. Roalter, and F. Michahelles, "Things that twitter: social networks and the internet of things.' What can the Internet of Things do for the Citizen (CIoT)," in *Proceedings of the 8th International Conference on Pervasive Computing (Pervasive 2010)*, 2010.

[19] R. Girau, S. Martis, and L. Atzori, "Lysis: A platform for iot distributed applications over socially connected objects," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 40–51, 2017.

[20] A. Pintus, D. Carboni, and A. Piras, "Paraimpu: A platform for a social Web of Things," in *Proceedings of the 21st Annual Conference on World Wide Web, WWW'12*, pp. 401–404, France, April 2012.

[21] I. Farris, R. Girau, L. Militano et al., "Social Virtual Objects in the Edge Cloud," *IEEE Cloud Computing*, vol. 2, no. 6, pp. 20–28, 2015.

[22] R. Girau, M. Nitti, and L. Atzori, "Implementation of an experimental platform for the social internet of things," in *Proceedings of the 7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2013*, pp. 500–505, July 2013.

[23] M. A. Jarwar, R. A. Abbasi, M. Mushtaq et al., "CommuniMents: A framework for detecting community based sentiments for events," *International Journal on Semantic Web and Information Systems*, vol. 13, no. 2, pp. 87–108, 2017.

[24] S. Newman, "Building microservices: designing fine-grained systems," 2015.

[25] "The Single Responsibility Principle - Programmer 97-things," 2017, http://programmer.97things.oreilly.com/wiki/index.php/The_Single_Responsibility_Principle.

[26] F. Viktor, "The DevOps 2.0 Toolkit: Automating the Continuous Deployment Pipeline with Containerized Microservices," 2016.

[27] "iCore: Internet Connected Objects for Reconfigurable Ecosystems, European FP7 Project," 2017, http://cordis.europa.eu/project/rcn/100873_en.html.

[28] D. Namiot and M. Sneps-Sneppe, "On microservices Architecture," *International Journal of Open Information Technologies*, vol. 2, no. 9, 2014.

[29] D. Bonino, M. T. D. Alizo, A. Alapetite et al., "ALMANAC: internet of things for smart cities," in *Proceedings of the 3rd*
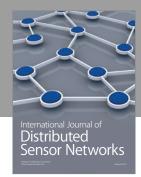
International Conference on Future Internet of Things and Cloud (FiCloud '15), pp. 309–316, IEEE, Rome, Italy, August 2015.

[30] "Developing Microservices for PaaS with Spring and Cloud Foundry," 2017, https://www.infoq.com/presentations/microservices-pass-spring-cloud-foundry.

[31] "Microservices in action, Part 2: Containers and microservices — a perfect pair," 2017, https://www.ibm.com/developerworks/cloud/library/cl-bluemix-microservices-in-action-part-2-trs/index.html.

[32] J. E. Kim, X. Fan, and D. Mosse, "Empowering End Users for Social Internet of Things," in *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation - IoTDI '17*, pp. 71–82, Pittsburgh, Pa, USA, April 2017.

[33] Z. U. Shamszaman, S. S. Ara, I. Chong, and Y. K. Jeong, "Web-of-Objects (WoO)-based context aware emergency fire management systems for the Internet of Things," *Sensors*, vol. 14, no. 2, pp. 2944–2966, 2014.

[34] S. Ali, M. G. Kibria, and I. Chong, "WoO enabled IoT service provisioning based on learning user preferences and situation," in *Proceedings of the 2017 International Conference on Information Networking (ICOIN)*, pp. 474–476, Da Nang, Vietnam, January 2017.

[35] M. A. Jarwar, S. Ali, M. G. Kibria, S. Kumar, and I. Chong, "Exploiting interoperable microservices in web objects enabled Internet of Things," in *Proceedings of the Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 49–54, Milan, Italy, July 2017.

[36] M. Kibria, S. Ali, M. Jarwar, S. Kumar, and I. Chong, "Logistic Model to Support Service Modularity for the Promotion of Reusability in a Web Objects-Enabled IoT Environment," *Sensors*, vol. 17, no. 10, p. 2180, 2017.

[37] M. Kibria, S. Fattah, K. Jeong, I. Chong, and Y. Jeong, "A User-Centric Knowledge Creation Model in a Web of Object-Enabled Internet of Things Environment," *Sensors*, vol. 15, no. 9, pp. 24054–24086, 2015.

[38] M. A. Jarwar and . Ilyoung Chong, "Exploiting IoT services by integrating emotion recognition in Web of Objects," in *Proceedings of the 2017 International Conference on Information Networking (ICOIN)*, pp. 54–56, Da Nang, Vietnam, January 2017.

[39] S. Ali, H.-S. Kim, and I. Chong, "Implementation model of WoO based smart assisted living IoT service," in *Proceedings of the International Conference on Information and Communication Technology Convergence, ICTC 2016*, pp. 816–818, October 2016.

[40] S. Kumar, M. G. Kibria, S. Ali, M. A. Jarwar, and I. Chong, "Smart spaces recommending service provisioning in WoO platform," in *Proceedings of the International Conference on Information and Communications (ICIC)*, pp. 311–313, Hanoi, Vietnam, June 2017.

[41] K. M. Alam, M. Saini, and A. El Saddik, "Toward social internet of vehicles: concept, architecture, and applications," *IEEE Access*, vol. 3, pp. 343–357, 2015.

[42] "Weka 3 - Data Mining with Open Source Machine Learning Software in Java," 2017, https://www.cs.waikato.ac.nz/ml/weka/.

[43] "Jena, Apache. &quot;A free and open source Java framework for building Semantic Web and Linked Data applications.&quot," 2017, http://jena.apache.org/.

[44] Protégé, "A free, open-source ontology editor and framework for building intelligent systems," 2017, https://protege.stanford.edu/.
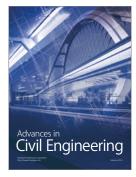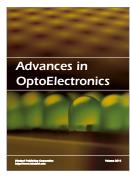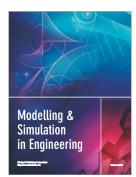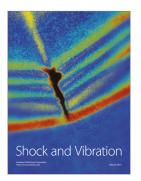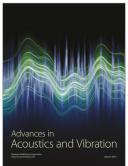
Multiple Antenna Technologies

Manar Mohaisen | YuPeng Wang | KyungHi Chang

The Graduate School of Information Technology and Telecommunications

INHA University

ABSTRACT

Multiple antenna technologies have received high attention in the last few decades for their capabilities to improve the overall system performance. Multiple-input multiple-output systems include a variety of techniques capable of not only increase the reliability of the communication but also impressively boost the channel capacity. In addition, smart antenna systems can increase the link quality and lead to appreciable interference reduction.

I. Introduction

Multiple antennas technologies proposed for communications systems have gained much attention in the last few years because of the huge gain they can introduce in the communication reliability and the channel capacity levels. Furthermore, multiple antenna systems can have a big contribution to reduce the interference both in the uplink and the downlink by employing smart antenna technology.

To increase the reliability of the communication systems, multiple antennas can be installed at the transmitter or/and at the receiver. Alamouti code is considered as the simplest transmit diversity scheme while the receive diversity includes maximum ratio, equal gain and selection combining methods. Recently, cooperative communication was deeply investigated as a mean of increasing the communication reliability by not only considering the mobile station as user but also as a base station (or relay station). The idea behind multiple antenna diversity is to supply the receiver by multiple versions of the same signal transmitted via independent channels.

On the other hand, multiple antenna systems can tremendously increase the channel capacity by sending independent signals from different transmit antennas. BLAST spatial multiplexing schemes are a good example of such category of multiple antenna technologies that boost the channel capacity.

In addition, smart antenna technique can significantly increase the data rate and improve the quality of wireless transmission, which is limited by interference, local scattering and multipath propagation. Through shaping the antenna radiation pattern and adaptively adjusting the antenna weight vector, smart antennas improve the communication link quality by increasing the received signal power and suppressing the interference.
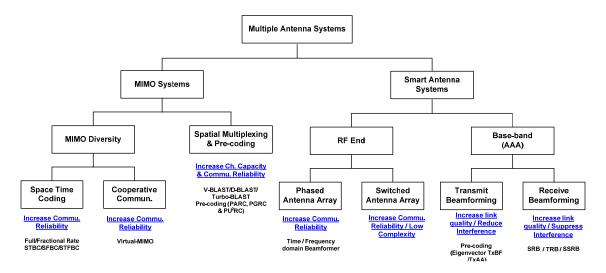
Fig. 1. Multiple antenna technologies.

Besides, on-line calibration technique is also adopted to correct the errors due to the distortions and nonlinearity of the radio frequency components in the antenna array system.

Fig. 1 summarizes the different multiple antenna technologies and gives some examples of these technologies.

This paper is organized as follows: in section II we present multiple antenna diversity schemes employed at the transmitter or/and at the receiver. Spatial multiplexing presented by BLAST schemes is detailed in section III. Section IV is dedicated to some advanced multiple input multiple-output (MIMO) systems including multi-user MIMO and cooperative communications. While techniques related to the smart antennas such as phased antenna array, switched beam antenna array, and adaptive antenna array are described in Section V. Finally, we conclude in Section VI.

II. **MIMO Diversity**

In communication systems, we have to increase the reliability of the communication operation between transmitter and receiver while maintaining a high spectral efficiency. The ultimate solution relies in the use of *diversity*, which can be viewed as a form of redundancy [1]. There are many diversity techniques that can be applied to communication systems; we mention herein time diversity, frequency diversity, and spatial diversity or any combination of these three diversities. In *time diversity*, the same information-bearing signal is transmitted in different time slots where a good gain can be achieved when the duration between the two slots, in which the same symbol is transmitted, is greater than the coherence time of the channel. In *frequency diversity*, the same information-bearing signal is transmitted on different subcarriers where a good diversity gain can be achieved when the separation between subcarriers is greater than the coherence bandwidth.

Finally, in *spatial diversity*, the same information-bearing signal is transmitted or received via different antennas where the maximum gain can be achieved when the fading occurring in the channel is independent (or low correlated). In the receiver, diversity gain can be achieved by combining the redundant signals arriving via independent (or lowly correlated) channels.

Fig. 2 shows some possible combinations of transmit diversity which can be achieved when employing multiple transmit antennas.

In the following section, we present some famous space-time block codes applied at the transmitter side. We present also the combining techniques used when different versions of the information-bearing signal are received. Finally, we present a scheme that includes transmit and receive diversities.

2.1 Space Diversity at the Transmit Side

The basic idea of the use of transmit diversity is to reduce the mobile station (MS) receiver complexity while improving the detection performance.

The pioneering work in the transmit diversity was done by Alamouti where he proposed his famous 2×1 space-time code. Alamouti scheme achieves diversity gain while requiring only a linear decoder.

Later on, Tarokh et al. proposed a generalized theory of the complex orthogonal space-time codes. Based on Tarokh work, more than two antennas can be used and the code rate can be fractional.

In the following we present the two different types of space-time codes.

2.1.1 Complex Orthogonal Space-Time Codes

For this type of space time codes, the following conditions must be satisfied

- Square transmission matrix (number of transmit antennas N_t equal to number of used time slots m)
- A unity code rate (number of used time slots m equals to number of transmitted symbols l)
- Orthogonality of the transmission matrix in the time and space domains ($SS^H = S^H S$) where S^H is the conjugate transpose of S.

Space-Time Diversity   **Space-Frequency Diversity**   **Space-Time-Frequency Diversity**
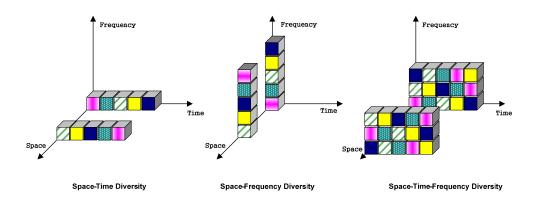
Fig. 2. Transmit diversity.
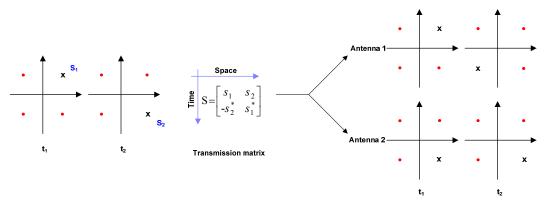
Fig. 3. Alamouti scheme example with QPSK modulation.

As said before, the simplest complex orthogonal space-time code is the Alamouti code which uses two transmit antennas and one receive antenna. Furthermore, Alamouti scheme requires that the fading channel envelope remains constant over two time slots.

Fig. 3 shows an example of the encoding process of Alamouti scheme with QPSK modulation [2], [3].

Fig. 4 shows the receiver structure used for decoding the combined received symbols.
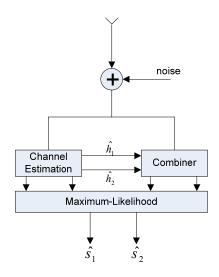
Fig. 4. Alamouti code receiver.

At the receiver the following signals are received (with applying the complex conjugate to the received signal at t_2)

$$\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} \qquad (1)$$

The linear combiner multiplies the received symbols by the Hermitian transpose of the channel matrix (for simplicity, we consider that channel is perfectly estimated). The output of the linear combiner is then given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_1^2 + h_2^2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \qquad (2)$$

Maximum-Likelihood (ML) decoder is then applied to get the transmitted symbols. As one can see, the simplicity of the receiver is due to the spatio-temporal orthogonality of the transmission matrix.

A complex orthogonal space-time code using 4 or 8 antennas was proposed by Tarokh et al. in [4].

2.1.2 Generalized Complex Orthogonal Space-Time Codes

The search for space-time codes with more than two antennas was started by Tarokh, Jafarkhani, and Calderbank. Their work

has built the basis for a theory of generalized complex orthogonal designs.

Generalized complex orthogonal designs are distinguished from Alamouti code by the following

- A non-square transmission matrix (number of used time slots ≠ number of Tx antennas)
- A fractional code rate (number of transmitted symbols < number of used time slots)
- Orthogonality of the transmission matrix is only guaranteed in the time sense.

As a consequence of these characteristics, the spectral efficiency is reduced and the number of time slots over which the channel should be constant is increased.

The transmission matrix of a generalized complex space-time code with 3 antennas, 4 transmitted symbols and 8 used time slots is given by [5]

$$
G_3 = \begin{bmatrix}
s_1 & s_2 & s_3 \\
-s_2 & s_1 & -s_4 \\
-s_3 & s_4 & s_1 \\
-s_4 & -s_3 & s_2 \\
s_1^* & s_2^* & s_3^* \\
-s_2^* & s_1^* & -s_4^* \\
-s_3^* & s_4^* & s_1^* \\
-s_4^* & -s_3^* & s_2^*
\end{bmatrix} \tag{3}
$$

In the literature, more research was done to increase the rate of the space-time codes. For more details refer to [6].

Table 1 summarizes the difference between Alamouti and space-time code characterized by the transmission matrix G_3.

These shown coding schemes can be transmitted in the space-time domains, space-frequency domains or in space-frequency-Time domains. These coding schemes are thus known as ST, SF, and STF coding, respectively [7].

2.1.3 Cyclic Delay Diversity (CDD)

CDD can be considered as a very simple transmit diversity scheme. CDD can achieve transmit artificial frequency diversity by selecting appropriate transmit delays. In this method, multiuser diversity, obtained by scheduling based on frequency domain channel response, can be improved by adjusting the delay spread (at the transmitter) which is done by controlling the delay values dependent on the channel condition [8], [9].

2.2 Space Diversity at the Receive Side

In space diversity at the receive side, multiple antennas are used in the receiver with sufficient spacing between antennas in such a way mutual correlation between antennas is reduced and as consequence diversity gain is increased [10]. To get diversity gain at the receiver, received signals from different antennas are combined. There are four combining methods, namely, select combining (SC), maximal-ratio combining (MRC), equal-gain combining (EGC), and square-law combining. The first three schemes are linear while the last requires a non-linear receiver.

Fig. 5 shows a simplified block diagram of the linear combining schemes which differ in the weighting vector **w.** In SC, the signal at the branch with maximum signal to noise ratio (SNR) is selected and other received signals are discarded.

The weighting vector $\mathbf{w} = (w_1, w_2, \ldots, w_M)$ is the Nth column of the identity matrix of size M where the Nth branch has the maximum SNR.

Table 1. Comparison between Alamouti and generalized complex space-time codes.

Space-time code	Number of Tx antennas	Number of transmitted symbols, l	Number of used time slots, m	Orthogonality of Tx matrix	Rate = l/m
S	2	2	2	Spatio-temporal sense	1
G_3	3	4	8	Only temporal sense	1/2

Table 2. Comparison between diversity combining schemes.

Scheme	Requiring CSI	Outage Probability $F(x)$	Application
SC	No	$\left[1-e^{-x/\bar{y}_c}\right]^M$	No constraints
MRC	Yes	$1-e^{-x/\bar{y}_c}\sum_{k=0}^{M-1}\frac{1}{k!}\left(\frac{x}{\bar{y}_c}\right)^k$	No constraints
EGC	Yes	No closed form for M > 2	No constraints
SLC	No	-	FSK or DS-CDMA

As one can see, the SC scheme does not require any channel information except that of SNR.   On the other hand, MRC and EGC schemes require the channel state information (CSI) or a part of it (channel envelope, phase, delay). MRC scheme weights the received signals according to their reliability; a more reliable signal has a high weight while a less reliable signal has a small weight. Also, the channel phase distortion is compensated. Finally, signals are aligned then combined. On the other hand, EGC scheme can be viewed as a simplified version of MRC where signals are weighted equally (i.e. the weighting vector $\mathbf{w} = [1, 1, …,1]_M$) then aligned before being combined coherently. In practice, the phase at different branches can't be often estimated. So, EGC and MRC can't be employed. In such situations, square law combining (SLC) can be applied to obtain spatial diversity without requiring phase estimation. Unlike linear combining schemes, SLC scheme can only be applied to modulation schemes which preserve some sort of orthogonality including frequency-shift keying (FSK) or direct-sequence CDMA [5], [11].

Table 2 summarizes a comparison between combining schemes. It is known that, from a performance point of view, MRC is optimum and gives the best performance among the pre-described combining schemes.
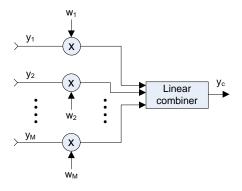
Fig. 5. Simple block diagram of linear combining schemes.

2.3 Combined Transmit/Receive Diversity

Spatial diversity schemes explained in the previous two sections can be combined together to achieve diversity at both receive and transmit sides. An example of such a hybrid spatial diversity scheme is

the 2×2 Alamouti/MRC MIMO system [12].

III. Spatial Multiplexing

As shown in the previous section, MIMO diversity can be used in the transmitter or the receiver sides or in both to increase the reliability of the communication. In this section we talk about spatial multiplexing schemes which are for goal to increase the channel capacity.

The most known spatial multiplexing schemes are the BLAST family which includes Vertical-BLAST, Diagonal-BLAST, and Turbo-BLAST. The acronym BLAST stands for "*Bell Laboratories Layered Space-Time*".

3.1 Diagonal-BLAST

D-BLAST was originally proposed by Foschini [13]. In D-BLAST, the symbols to be transmitted are arranged on the diagonals of the space-time transmission matrix where elements under the diagonal are padded with zeros. Fig. 6-a depicts the structure of the D-BLAST transmitter for four transmit antennas. At first, the bit stream is de-multiplexed into four parallel streams which are encoded and modulated independently. Encoded-modulated streams are cycled over time. Equation (4) is an example of the transmission matrix when using four transmit antennas.

$$S = \begin{bmatrix} s_{1,1} & s_{1,2} & s_{1,3} & s_{1,4} & \cdots & s_{1,K-1} & s_{1,K} & 0 & 0 & 0 \\ 0 & s_{2,1} & s_{2,3} & s_{2,4} & \cdots & s_{2,K-2} & s_{2,K-1} & s_{2,K} & 0 & 0 \\ 0 & 0 & s_{3,1} & s_{3,2} & \cdots & s_{3,K-3} & s_{3,K-2} & s_{3,K-1} & s_{3,K} & 0 \\ 0 & 0 & 0 & s_{4,1} & \cdots & s_{4,K-4} & s_{4,K-3} & s_{4,K-2} & s_{4,K-1} & s_{4,K} \end{bmatrix} \quad (4)$$

(a) D-BLAST

(b) V-BLAST

(c) Turbo-BLAST

(d) RLST code
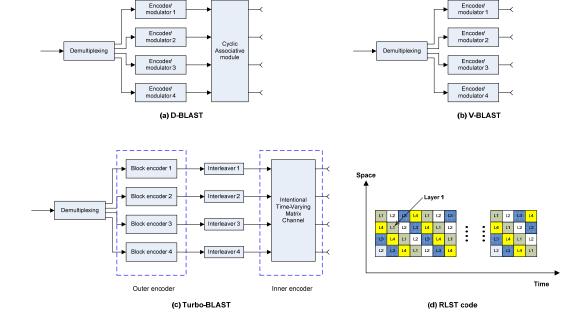
Fig. 6. Transmitter block diagrams for BLAST family using four transmit antennas.

Table 3. MIMO systems diversity orders.

MIMO Configuration	Diversity order
STBC	$N_t \times N_r$
BLAST	$N_r - N_t + 1$

The first diagonal of S is transmitted via the first antenna; the second diagonal is transmitted via antenna 2, and so on.

3.2 Vertical-BLAST

A simplified version of D-BLAST was proposed by Wolniansky known as Vertical-BLAST or V-BLAST [14]. In V-BLAST, incoming data stream is de-multiplexed into N_t streams each of which is encoded and modulated independently and sent on an antenna of its own. V-BLAST high-level diagram is depicted in Fig. 6-b where four antennas are used at the transmit side. Compared to D-BLAST, V-BLAST does not include cycling over time, the complexity is significantly reduced. In addition, unlike D-BLAST, V-BLAST does not include any space-time wastage. At the receiver, transmitted symbols can be decoded using *ordered serial interference-cancellation (OSIC) detector*. For the OSIC to work properly, the number of receive antennas N_r must be at least as large as the number of transmit antennas.

3.3 Turbo-BLAST

Turbo-BLAST was first described by Sellathurai and Haykin [15]. The Turbo-BLAST transmitter structure is depicted in Fig. 6-c. The data stream bits are firstly demultiplexed into N_t parallel streams which are encoded independently using the block encoder (outer encoder) (i.e. channel coding). The output streams of the outer encoder are interleaved independently and passed to the inner encoder. The mission of the outer encoder is to achieve *random-layered space-time* (RLST) coding.

The structure of the RLST encoder, with periodical cyclic space-time interleaving is depicted in Fig. 6-d. For optimal performance of the RLST code, the receiver should employ the *maximum a posteriori probability (MAP)* decoding algorithm. Nevertheless, the complexity of the MAP decoding algorithm is very high (increases exponentially with N_t). To decrease the complexity of the receiver, the near-optimal turbo-like receiver can be used. This near-optimal turbo-like receiver is known as *iterative detection and decoding (IDD) receiver*.

Before going further, we list in Table 3 a comparison between diversity order of the different space-time coding and the BLAST family schemes.

IV. Advanced Topics

4.1 Single and Multi CodeWord MIMO

In single codeword (SCW) MIMO, an encoded packet is distributed across many streams to form the MIMO transmission. Feedback is used to control the rank of the MIMO transmission (number of streams used) as well as the overall rate of transmission. In multiple codeword MIMO, several separately encoded packets are transmitted independently over the multiple streams. Here the rate of each stream can be controlled with feedback [16] and [17].

4.2 Single-User MIMO and Multi-User MIMO

In single-user MIMO, already explained techniques in previous sections are used where the channel capacity grows linearly with *min(Nt, Nr)* [18].

For multi-user MIMO, which is of high interest research topic, it was shown that for N_t transmitting antennas (at the base station) and N_r users, the same overall capacity can be achieved. This later work was encouraged by applying dirty paper coding [19] where results showed that if the transmitter knows the interfering signal, then the channel capacity will not be affected by the presence of the interference [20]. On the other hand, multi-user MIMO can integrate beamforming to apply spatial division multiple access (SDMA).

4.3 Cooperative Communication and Virtual MIMO

In cooperative communication, a mobile can act as both a user and relay. As consequence, mobile sends to the base station its own data bits and some of other mobiles (sometimes called partner) information bits. Fig. 7 shows a cooperative cellular system where for simplicity we consider three cooperative users and one base station [21]. As depicted in Fig. 7, user 1 cooperates with users 2 and 3 to send its own information. As a result, the overall cooperative system can be seen as virtual-MIMO (V-MIMO) and in the above example it is $3 \times N_{BS}$ MIMO system (for the uplink) where N_{BS} is the base station's number of receive antennas. Users 2 and 3 can simply amplify and forward user 1 received information or detect and forward [22]. Another method of cooperation is the coded cooperation where different coded portions are sent via different fading channels [23].
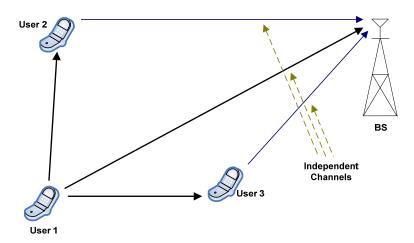
Fig. 7. Cooperative communication and virtual-MIMO.

4.4 Pre-Coded MIMO with Rank Adaptation

4.4.1 Per Antenna Rate Control (PARC)

PARC can be considered as a closed-loop MIMO system where transmitter uses *channel quality indication* (CQI) fed by the receiver to select the best modulation and coding schemes per antenna. Fig. (8) shows a general PARC transmitter structure with 4 transmit antennas [24].

4.4.2 Per Group Rate Control (PGRC)

In PARC, a CQI feedback is necessary for each transmit antenna. This increases the uplink overhead.
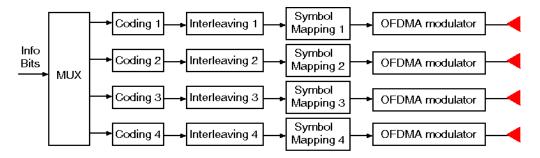
Fig. 8. Per Antenna Rate Control (PARC).

To solve this problem, PGRC is used where a feedback is required per group of antennas. This reduces the feedback information while maintaining almost the same performance of PARC [24].

4.4.3 Per User Unitary Rate Control (PU^{2}RC)

PU^{2}RC is a multi-user closed-loop MIMO system. Each user feeds back the CQI to the base station. The base station uses the CQI to determine the modulation and coding schemes per user. In addition, base station can apply unitary pre-coding and adaptively select the number of transmit antennas (rank adaption) [25].

V. Smart Antennas

5.1 Introduction

Smart antenna was born in the early 1990s when well developed adaptive antenna arrays originate from Radar system. Later, Smart antenna technique is applied in wireless communications system. Recently, Smart antenna technique has been proposed as a promising solution to the future generations of wireless communication systems, such as the Fourth-Generation mobile communication systems, broadband wireless access networks, where a wide variety of services through reliable high-data rate wireless channels are expected. Smart antenna technique can significantly increase the data rate and improve the quality of

wireless transmission, which is limited by interference, local scattering and multipath propagation [26], [27]. Smart antennas offer the following main applications in high data-rate wireless communication systems [28], [29]:

- Spatial Diversity
- Co-channel interference reduction
- Angle reuse or space division multiple access (SDMA)
- Spatial multiplexing

Smart antenna system can be categorized into three main groups: Phased antenna array system, switched beam systems, and adaptive antenna array system. To match the characteristics in each radio frequency chain of the transmitter and receiver, on-line calibration is required in smart antenna systems. On-line calibration technique can compensate the errors such as the distortions of radio frequency components due to small environment changes, the nonlinear characteristics of mixer, amplifier and attenuator, I/Q imbalance errors, etc.

5.2 Phased Antenna Array System

Phased antenna array is a group of antennas in which the relative phases of the respective signals feeding the antennas are varied in such a way that the effective radiation pattern of the array is reinforced in a desired direction and suppressed in undesired directions.

Phased antenna array system is usually utilized in *radio frequency* (RF) or *intermediate frequency* (IF) with the system central frequency larger than 10 GHz, such as satellite communication system [30]. There are two main different types of phased arrays, also called beamformers. There are time domain beamformers and frequency domain beamformers.

5.3 Switched Beam System

The switched beam method is considered as an extension of the current sectorization scheme. In the switched beam approach, the sector coverage is achieved by multiple predetermined fixed beam patterns with the greater gain placed in the centre of a beam [30]. When a mobile user is in the vicinity of a beam, then the signals at the output ports will be given as in (5). This enables the switched beam system to select the signal from the output port corresponding to that beam. As the mobile moves to the coverage of another beam during the call, the system monitors the signal strength and switches to other output ports as required. A basic switched beam antenna architecture is shown in Fig. 9. And Fig. 10 illustrates the produced antenna pattern with 4 antennas.

$$y_i(t) = s_i(t)G_i(t)\sum_{l=1}^{L} I_l(t)G_{li}(\theta_l) \quad (5)$$

where $y_i(t)$ is the total signal appearing at port i, $s_i(t)$ is the signal source, $I_l(t)$ is the interfering signal source located at arbitrary angles θ_l, G_i is the transfer function between signal source along the main beams and their corresponding output ports, G_{li} is the transfer function between interference signal l and port i.
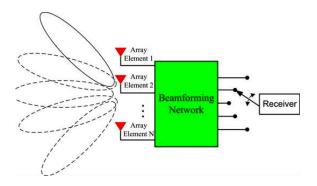
Fig.9. Functional block diagram of switched beam system.
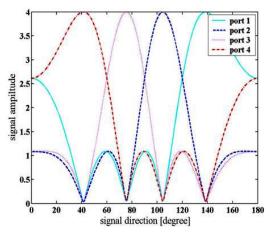
Fig.10. Produced antenna pattern of switched beam system with 4 antennas.

Switched beam systems can offer several advantages, including

- Low complexity and cost. Since switched beam system only requires a beamforming network, RF switches, and simple control logic, they are relatively easy and cheap to implement.
- Moderate interaction with base station receivers. In practice, switched beam system can simply replace conventional sector antennas without requiring significant modifications to the radio base station antenna interface or the baseband algorithms implemented at the receiver.
- Coverage extension. The antenna array aperture gain will boost the link

budget, which could be translated to a coverage extension.

5.4 Adaptive Antenna Array System (AAA)

Adaptive antennas date back to 1959. The original work was attributed to L. C. Van Atta's work, Electromagnetic Reflection. Since then, adaptive beamforming techniques have been employed to remove unwanted noise and jamming from the output, mainly in military applications. With the thriving commercial wireless communication industry and the advancing microprocessor technologies, the adaptive beamforming techniques have found their applications in commercial wireless communications. With powerful *digital signal processing* (DSP) hardware at the base-band, algorithms could control antenna beam patterns adaptively to the real signal environment, forming beams towards the desired signals while forming nulls to co-channel interferers. Thus, the system performance is optimized in terms of link quality and system capacity [31]. Adaptive antenna array can be utilized in the transmitter side, which is known as *transmit beamforming* (TxBF) or in the receiver side, which is called *receive beamforming* (RxBF).

5.4.1 Transmit Beamforming (TxBF)

The implementation of adaptive antenna array technique in a handset is difficult with today`s hardware due to its limitations in size, cost, and energy storage capability, while it is feasible to adopt antenna arrays at base stations.

Transmit beamforming provides a powerful method for increasing downlink capacity [32]-[35]. The idea of TxBF is similar to the pre-coded MIMO technique but with different strategies to calculate the transmit weight vector. TxBF adjusts the antenna main lobe towards to the desired user and reduce the interference to other users. A simple illustration of TxBF is shown in Fig. 11.
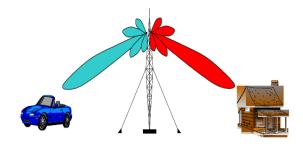
Fig.11. An illustration of TxBF.

Eigenvector TxBF Algorithm

Eignenvector TxBF algorithm is widely used for TxBF. The eigenvector of the spatial covariance channel matrix is calculated as

$$\mathbf{R}_{ss} = \lambda \mathbf{H} \qquad (6)$$

where $\mathbf{R}_{ss}$ is the autocovariance matrix of the desired user`s signal, and $\mathbf{H}$ is the spatial covariance channel matrix. The eigenvector λ_{max} which corresponds to the largest eigenvalue will be selected as the weight vector [36]. One example of beam pattern for 4 uniform linear array elements is shown in Fig 12.
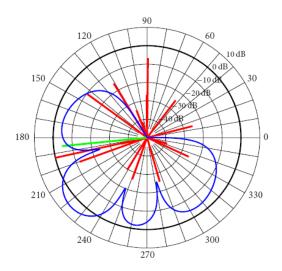
Fig. 12. Example beam pattern of 4 antenna elements in a sectorized system for a single sector (main beam direction is 240°).

Transmit Adaptive Array (TxAA) Algorithm

Transmit adaptive array (TXAA) is a technique in which the user periodically sends quantized estimates of the optimal transmit weights to the BS via a feedback channel. The transmitter weights are optimized to deliver maximum power to the user. The optimal transmit weights are given by

$$\mathbf{w} = \mathbf{H}^H / \mathbf{H}\mathbf{H}^H \qquad (7)$$

where $\mathbf{w}$ is the transmit weight vector and $\mathbf{H}$ is the channel matrix.

The weights are normalized so that the total transmitted power is not altered. In the case of multipath channels emanating from each antenna, the optimal weights will be given by the principal eigenvector of the channel correlation matrix $\mathbf{H}^H \mathbf{H}$.

5.4.2 Receive Beamforming (RxBF)

Beamforming also can be applied in the uplink to improve the link quality and suppress the co-channel interference, which is known as *receive beamforming*

(RxBF). Through RxBF, smart antenna system can receive predominantly from a desired direction (direction of the desired source) compared to some undesired directions (direction of interfering sources). This implies that the digital processing has the ability to shape the radiation pattern to adaptively steer beams in the direction of the desired signals and put nulls in the direction of the interfering signals. This enable low co-channel interference and large antenna gain to the desired signal.

Based on the reference signals adopted in the beamforming algorithms, RxBF can be classified into *spatial reference beamforming* (SRB), *temporal reference beamforming* (TRB), and *signal structure reference beamforming* (SSRB).

Spatial Reference Beamforming (SRB)

Spatial reference beamforming method is sometimes referred as *direction of arrival* (DoA) method. SRB estimates the direction of arrival of the signal based on the spatial reference signal, using any of the techniques like *multiple signal classification* or *estimation of signal*
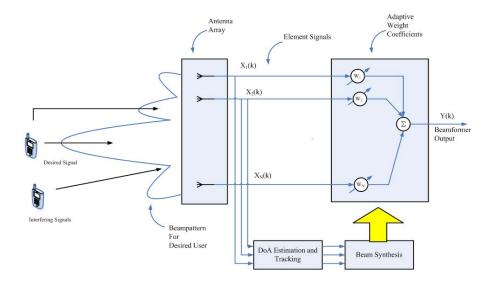
Fig. 13. A general structure of SRB.

parameters via rotational invariance techniques algorithms or their derivatives. They involve finding a spatial spectrum of the antenna/sensor array, and calculating the DoA from the peaks of this spectrum [37]. A general architecture of SRB algorithm is shown in Fig. 13. The general steps of SRB method are shown as follows:

- DoA Estimation
  - Arbitrary Array: MUSIC, etc.
  - Linear Array: ESPRIIT, etc.
- Beam Synthesis
  - Gram-Schmidt, etc.
- Combining
  - EGC, MRC, Wiener Filter, etc.

Multiple signal classification (MUSIC) algorithm estimates the DoA of the desired signal by using an eigen-space method based on a spatial reference signal. MUSIC requires intensive calculation of eigenvalues and eigenvectors of an autocorrelation matrix of the input vectors from the receiving antenna array. A general step of MUSIC algorithm is shown below:

- Collect received samples and estimate the covariance matrix of the received samples.
- Perform eigen-decomposition of the covariance matrix.
- Calculate spatial spectrum.
- Estimate DoA by locating peaks in the spectrum.

Estimation of signal parameters via rotational invariance techniques (ESPRIT) is also well known for the SRB method. In addition, ESPRIT has many important advantages over MUSIC algorithm [38]:

- No knowledge of the array geometry and element characteristics are required.
- Much less complex on computation.
- No calibration of the array is required.
- The algorithm simultaneously estimates the number of sources and DoA`s
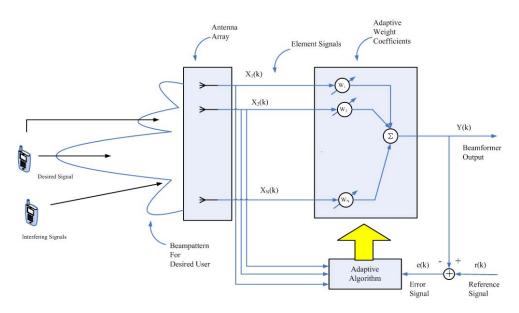
Fig. 14. A general structure of TRB.

Temporal Reference Beamforming (TRB)

Temporal Reference Beamforming shown in Fig. 14, is a method used to create the radiation patter of the antenna array by adding constructively the phases of the signals in the DoA of the desired user, and nulling the pattern of the interfering users based on the temporal reference signal [39]. Based on the temporal reference signal and some predefined adaptive weight calculation criterion, some adaptive algorithms such as LMS (Least Mean Square), RLS (Recursive Least Squares), and DSMI (Direct Sample Matrix Inverse) algorithms, are used to adjust the weight vector of the antenna array to improve the link quality. The general characteristics of TRB are as follows:

- Good performance in multipath channel environment
- Computationally inexpensive
- Requiring Training sequence
- Difficult to apply TxBF because of the absence of DoA information

Criterion of Adaptive Weight Calculation

In the *minimum mean-square error* (MMSE) criterion, the weights are chosen to minimize the *mean-square error* (MSE) between the beamformer output and the temporal reference signal. While in the *maximum signal-to-interference ratio* (MSIR) criterion, the weights are chosen to directly maximize the signal-to-interference ratio (SIR). And The *minimum variance* (MV) criterion chooses the weights that minimize the variance of the output power. All the above three criterions has the same form of

$$\mathbf{w}_{opt} = \beta \mathbf{R}_i^{-1} \mathbf{v} \qquad (8)$$

where $\mathbf{R}_i^{-1}$ is the inverse of the covariance matrix of the interference signals received in the antenna array and $\mathbf{V}$ is the antenna array propagation vector [40].

Let us assume that *d(t)* is the transmitted temporal reference signal and $\mathbf{R}_u$ is the covariance matrix of interference signals at the output of the beamformer. The calculation of β for MMSE, MSIR

Table 4. β Calculation.

Criterion	MMSE	MSIR	MV
β	$\dfrac{E\{d^2(t)\}}{1+E\{d^2(t)\}\mathbf{v}^H\mathbf{R}_i^{-1}\mathbf{v}}$	$\dfrac{E\{d^2(t)\}}{SIR}\mathbf{v}^H\mathbf{w}_{opt}$	$\dfrac{g}{\mathbf{v}^H\mathbf{R}_u^{-1}\mathbf{v}}$

and MV criterions are summarized in Table 4.

Adaptive Beamforming Algorithm

The *least mean square* (LMS) algorithm uses the temporal reference signal to update the weights at each iteration. In the LMS algorithm, we are searching for the optimal weight that would make the array output either equal or as close as possible to the reference signal, which is the weight that minimizes the MSE. Since the MSE has a quadratic form, moving the weights in the negative direction of the gradient of the MSE should lead us to the minimum of the error surface. The weight update equation is shown in (9) [30].

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \mu\mathbf{x}(t+1)\varepsilon^* \quad (9)$$

where μ is a constant, called the step size, which determines how close the weights approach the optimum value after each iteration and it controls the convergence speed of the algorithm. And ε is the error signal between the temporal reference signal and the received signal at the beamformer output. $\mathbf{x}(t+1)$ is the received signal vector at the antenna array at time $t+1$.

The main drawback of the LMS algorithm is that it is sensitive to the scaling of its input. This makes it very hard (if not impossible) to choose a step size μ that guarantees stability of the algorithm. The *normalized least mean square* (NLMS) algorithm is a variant of the LMS algorithm that solves this problem by normalizing with the power of the input. The weights updating function of NLMS algorithm is shown as

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \frac{\hat{\mu}}{a + \|x(t)\|^2}\mathbf{x}(t)\varepsilon^* \quad (10)$$

Recursive least square (RLS) algorithm is derived to overcome the drawback of slow convergence speed in the LMS algorithm, when the eigenvalue spread of the correlation matrix R of received signal vector $\mathbf{x}$ is large. RLS algorithm replaces the step size μ with the inverse of R. The weights are then updated using (11).

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \mathbf{R}^{-1}\mathbf{x}(t+1)\varepsilon^* \quad (11)$$
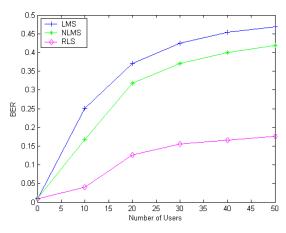
Fig. 15. Performance comparison among LMS, NLMS, and RLS algorithms.

Fig. 15 shows a simple performance comparison of the above three algorithms in OFDMA system under the Rayleigh fading channel with 8 antenna elements [41]. From this figure, we see that RLS algorithm performs best due to its faster

convergence speed than LMS and NLMS algorithms.

Signal Structure Reference Beamforming (SSRB)

SSRB method is based on inherent structure of the transmit signal of the implicit kind reference signal. Algorithms such as blind beamforming, least squares, and constant modulus algorithms, are based on the SSRB method. SSRB method is robust against different propagation conditions and does not require the array manifold knowledge. But the convergence problem becomes the main drawback of the SSRB method.

VI. Conclusions

In this paper we introduced the multi antenna technologies which can be considered as one of the most vivid area of research. Multiple antenna technologies were categorized into two main groups where in the first group we introduced some techniques related to spatial diversity and spatial multiplexing by outlining the gain achieved by these schemes. Furthermore, we introduced the smart antenna techniques and the up-to-date research progress in this field.

The advantages of multiple antenna systems make of them a very strong candidate to increase link reliability, increase channel capacity and reduce interference in both uplink and downlink.

References

[1] A. Goldsmith, *Wireless Communications*, Cambridge Univ. Press, 2005.

[2] M. Alamouti, "A simple transmit diversity technique for wireless communications", *IEEE J. Sel. Area Commun.,* 16, pp. 1451-1458, 1998.

[3] G. Tsoulos, *MIMO System Technology for Wireless Communications*, Taylor and Francis, 2006.

[4] V. Tarokh et al., "Space-time block codes from orthogonal designs", *IEEE Trans. Inf. Theory,* vol. 45, no. 5, pp. 1456-1467, 1999.

[5] S. Haykin and M. Mohr, *Modern Wireless Communications,* Pearson Prentice Hall, 2005.

[6] W. Su and X-G. Xi, "Two generalized complex orthogonal space-time block codes of rates 7/11 and 3/5 for 5 and 6 transmit antennas", *IEEE Trans. Inf. Theory,* vol. 49, no. 1, pp. 313-316, Jan 2003.

[7] K. Suto, and T. Ohtsuki, "Space-time-frequency block codes over frequency selective fading channels," *IEICE Trans. Commun.,* vol. E87-B, no. 7, pp. 1939-45, July 2004.

[8] NTT DoCoMo, "Multi-degree cyclic delay diversity with frequency-domain channel dependent scheduling", *3GPP TSG-RAN WG1 meeting #44bis*, R1-060991, Mar. 2003.

[9] Samsung, "Adaptive cyclic delay diversity," *3GPP TSG-RAN1 43#*, R1-051354, 7th – 11st, Nov. 2005, Seoul, Korea.

[10] G. Stüber, *Principles of Mobile Comm-unication,* Kluwer, 2001.

[11] M. K. Simon and M-S. Alouini, *Digital Communication over Fading Channels*, Wiley, 2005.

[12] E. Biglieri et al., "Diversity, interference cancellation and spatial

multiplexing in MIMO mobile WiMAX systems", *to appear in IEEE WiMAX '07*, 2007.

[13] G. Foschini, "Layered space-time architecture for wireless communication in fading environment when using multi-element antennas", *Bell Labs Technical Journal,* Autumn 1996.

[14] P. Wolniansky et al., "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel", *URSI International Symposium on Signals,* Systems and Electronics, 1998.

[15] M. Sellathurai and S. Haykin, "TURBO-BLAST for wireless communications: theory and experiments", *IEEE Trans. Signal Processing,* vol. 50, no., 10 pp. 2538-2546, Oct. 2002.

[16] A. Hottinen et al., "Industrial embrace of smart antennas MIMO," *IEEE Wireless Commun. Mag.*, Aug. 2006.

[17] A. Jette et al. "UMBFDD candidate proposal for IEEE 802.20," IEEE C802.20-07/09, Mar. 2007.

[18] A. Goldsmith et al., "Capacity limits of MIMO channels," *IEEE J. on Sel. Areas in Commun.*, vol. 21, no. 5, pp. 684-702, Jun. 2003.

[19] M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[20] Q. Spencer et al., "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.,* pp. 60-67, Oct. 2004.

[21] A. Nosratinia and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Communications Magazine,* pp. 74-80, Oct. 2004.

[22] J. Laneman et al., "An efficient protocol for realizing cooperative diversity in wireless networks," *Proc. IEEE ISIT*, June 2001, pp.294.

[23] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. on Wireless Commun.,* vol. 5, no. 2, pp. 1-7, Feb. 2006.

[24] Texas Instruments, "MIMO OFDMA E-UTRA proposal for different antenna configurations," *3GPP TSG RAN1 WG1 #43*, R1-051315, Nov. 2005.

[25] Samsung, "Downlink MIMO for EUTRA," *3GPP TSG RAN1 WG1 #43*, R1-051353, Nov. 2005.

[26] S. Ohmori, Y. Yamao, and N. Nakajima, "The future generationsof mobile communications based on broadband access technologies," *IEEE Commun. Mag.,* vol. 38, pp. 134–142, Dec. 2000.

[27] K. Sheikh, D. Gesbert, D. Gore, and A. Paulraj, "Smart antennas for broadband wireless access networks," *IEEE Commun. Mag.,* vol. 37, pp. 134–142, Nov. 1999.

[28] A. Lozano, F.R. Farrokhi, and R.A. Valenzuela, "Lifting the limits on high-speed wireless data access using antenna arrays," *IEEE Commun. Mag.,* pp. 156–162, Sept. 2001.

[29] Jeffrey H. Reed," Smart antennas: A system level overview for software defined radios for creating an API", SDRF-04-I-0057-V0.00, Software Defined Radio Forum, Jan. 2004.

[30] Ahmed EI Zooghby, *Smart Antenna Engineering*, Artech House, 2005.

[31] Michael Chryssomallis, "Smart antennas", *IEEE Antennas and Propagation Mag.,* vol. 42, pp.129-136, , June 2000.

[32] H. Boche and M. Schubert, "Theoretical and experimental comparison of optimization criteria for downlink beamforming," *European*

Trans. on Telecomm., vol. 12, no. 5, pp. 417–426, 2001.

[33] R. M. Buehrer, A. G. Kogiantis, S.-C. Liu, J. Tsai, and D. Uptegrove, "Intelligent antennas for wireless communications -uplink," *Bell Labs Technical Journal,* vol. 4, no. 3, pp. 73–103, 1999.

[34] H. Holma and A. Toskala, *WCDMA for UMTS*, John Wiley & Sons, 2000.

[35] A. Yener, R. D. Yates, and S. Ulukus, "Interference management for CDMA systems through power control, multiuser detection, and beamforming," *IEEE Trans. on Comm.,* vol. 49, no. 7, pp. 1227–1239, 2001.

[36] S.J. Ko, J. Heo, and K.H. Chang, "An effective downlink resource allocation for supporting heterogeneous traffic data in an OFDM/SDMA-based cellular system," *in Proc. IEEE GLOBECOM*, Nov. 2006, WLC27-5.

[37] Lal C. Godara, "Application of antenna arrays to mobile communications, part II: beam-forming and direction-of-arrival considerations", *IEEE Proc.,* vol .85, pp.1195-1245, Aug. 1997.

[38] Haardt, M., Nossek, J.A., "Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden," *IEEE Trans. on Signal Processing,* vol.43, pp. 1232 -1242, 1995.

[39] B. L. P. Cheung, Simulation of Adaptive Array Algorithms for OFDM and Adaptive Vector OFDM Systems, Master of Science in Electrical Engineering of Virginia Polytech., Sept. 2002.

[40] John Litva and Titus Kwok-Yeung Lo, *Beamforming in Wireless Communication*, Artech House, 1996.

[41] J. Heo and K.H. Chang, "Transmit and receive beamforming for OFDMA/TDD system," *in Proc. ISAP*, Aug. 2005, pp.27-30.

Manar Mohaisen

July, 2001: BS, Communication and Control, University of Gaza, Gaza, Palestine

Sep., 2005: MS，The School Polytechnic of Nice University, Sophia-Antipolis, France

Feb., 2006 ~ Present: Ph.D. student, The Graduate School of IT & T, INHA University

2001 ~ 2003: The Palestinian Telecommunication Company (JAWWAL)

<Rresearch Interests> MIMO Detection and Co-Channel Interference Cancellation

Yupeng Wang

July, 2004: BS, Communication Engineering, Northeastern University, Shenyang, China

July, 2006: MS, The Graduate School of IT & T, INHA University

Sep., 2006 ~ Present: Ph.D. student, The Graduate School of IT & T, INHA University

<Research Interests> 3GPP LTE Systems, Radio Resource Management, MIMO Techniques, and UWB

KyungHi Chang

Feb., 1985: BS, Electronics Engineering, Yonsei University

Feb., 1987: MS, Electronics Engineering, Yonsei University

Aug., 1992: Ph.D., EE Dept., Texas A&M Univ.

1989 ~ 1990: A Member of Research Staff, Samsung Advanced Institute of Technology (SAIT)

1992 ~ 2003: A Principal Member of Technical Staff (Team Leader), Electronics and Telecommunications Research Institute (ETRI)
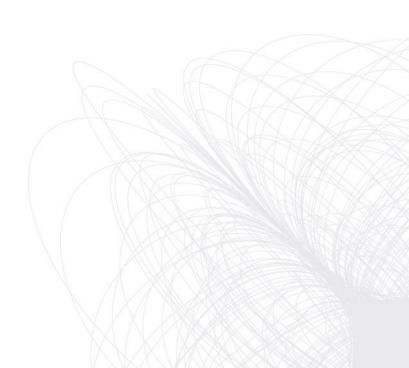
2003 ~ Present: Associate Professor, The Graduate School of IT & T, INHA University

<Research Interests> RTT design for IMT-Advanced & 3GPP LTE Systems, WMAN System Design, Cognitive Radio, Cross-layer Design, and Cooperative Relaying System

arcep
autorité de régulation
des communications électroniques
et des postes

5G: ISSUES & CHALLENGES

March 2017

The issues and challenges surrounding 5G

The telecoms industry is currently in the process of designing the technologies that are due to take over from 4G, which is still being deployed today. A great deal of work is thus underway to prepare these "5G" technologies.

To prepare for the arrival of this new generation of technologies, Arcep wanted to take a detailed look at the industry to better understand what is in the works. This report is the fruit of the interviews and research that Arcep conducted over the course of 2016, and which the Authority wanted to publish as a way to contribute to the public debate over 5G.

Its aim it to provide as objective and exhaustive an overview as possible, and deliver a concise, informative snapshot of the work that is currently underway on the future generation of mobile networks.

This report reflects the views of the stakeholders who were interviewed, but in no way represents Arcep's positions on or roadmap for 5G. Arcep awarded in 2015 the 700 MHz band and is currently working towards licensing the 3.5 GHz band, both of which have been identified as 5G bands. Arcep is also working with the Direction générale des entreprises and the Agence nationale des fréquences towards enabling spectrum for 5G.

Arcep would like to thank all of the entities (listed on the last page) who agreed to take part in this process, and who were willing to contribute to the regulator's investigation into the development of 4G's successor. Arcep is dedicated to sustaining an ongoing dialogue with the market, and invites all interested parties to provide information on the matter or to share their reactions to this report, by contacting Arcep's Mobile and Innovation Directorate at the following address: 5G@arcep.fr.

Table of Contents

Introduction: the vision for 5G, creating an ultra-connected society

5G wants to be the disruptive generation, the generation that no longer only caters to the needs of mobile operators and consumer communications, but which opens up new prospects and enables an extremely wide diversity of applications and use, unified within a single technology. 5G is setting itself up as an enabler of the digitisation of society and the economy.

The idea that is starting to take shape behind the notion of 5G is that it will not mean just an increase in transmission speeds, as has been the case with previous generations.

Consumer mobile communications, video downloads and the use of mobile apps account for the bulk of 4G networks' radio resources usage today. With 5G, the goal will be to enable a much broader spectrum of uses and a much greater diversity of users.

5G is targeting a wide variety of sectors, which will not necessarily have anything other than this technology in common, but which are central pillars in a society: energy, healthcare, media, industry and transportation.
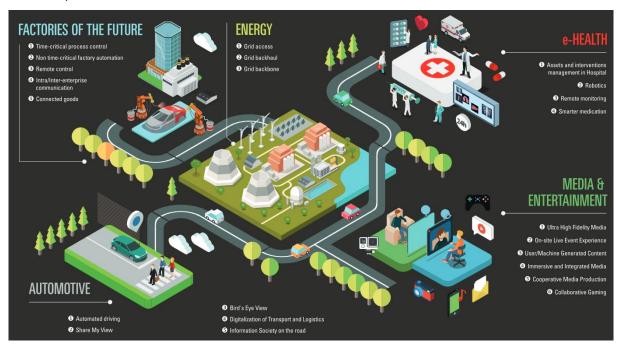
Figure 1. 5G driving industrial and societal changes[1]

The energy sector, for instance, has undergone a great many changes and developments over the past several years in terms of energy production, storage and transport. Rising fossil fuel costs, the introduction of renewable energy sources and markets being opened up to competition have all helped usher in new kinds of energy product and new stakeholders – both independent companies and ordinary citizens – to the energy marketplace. The objective with 5G will thus be to enable better management of these networks (smart grids, smart agriculture, factories of the future) and their interconnections, to achieve more efficient and more agile distribution.

---

[1]5G Empowering vertical industries. White Paper, 2016, https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf

Healthcare, transportation and media are sectors that are bound up with our fellow citizens' daily lives: any improvements in these sectors thus have a clearly visible impact. The introduction of robots to perform specialised surgical operations, of video on demand or the advent of new connected features in cars have already improved our daily lives. 5G is promising to go several steps further in all of these areas: it would enable remote medical diagnoses and operations in real time, it would democratise streaming of 360° 3D video, it would provide users with a vast selection of video content with a picture quality better than ultra high-definition (4K, 8K…). The automotive universe could rely on these new networks to help cars make decisions without human involvement, and also communicate with one another (this is already possible, for instance, with the first experimental fleet of fully autonomous taxis being tested on the streets of Singapore by the firms nuTonomy and Grab[2]), with reaction times that are compatible with the demands of high-speed travel.

In terms of factories of the future, the improvements brought by 5G are primarily targeting the introduction of new generations of connected robots, the interconnection of production sites and the much heavier use of smart sensors to improve industrial processes. Generally speaking, the aim is to achieve ubiquitous communication between machines, a process that is already well underway.

One thing that it is crucial to understand, through the few examples listed above whose demands might seem mutually incompatible, is that depending on the sector or the application, the network properties and the functionalities required will not be the same. Service providers – whether they are today's mobile operators or other market players – will need to be capable of adapting their network to demand, occasionally in real time. 5G will thus be not so much a universal technology as a polymorphous, or multi-faceted technology, capable of adapting to any use, up to and including the most demanding ones.

---

[2] http://nutonomy.com/
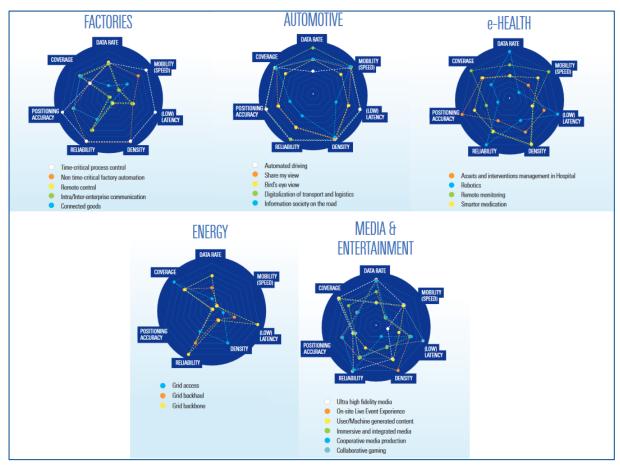
Figure 2. Performances required by vertical sectors[3]

---

[3] 5G Empowering vertical industries. White Paper, 2016, https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf

1   5G objectives

Ever since the first real mobile telephone call, 44 years ago[4], mobile technologies have been evolving steadily and their performances have improved exponentially: as mentioned earlier, calling services and later texting and finally the mobile internet and the use of multi-service mobile applications have characterised the evolution of mobile networks and the transition from one generation to the next. The birth of LTE technology and the fourth generation (4G), coupled with the widespread use of smartphones and tablets, have driven a massive increase in the amount of mobile data traffic being relayed over the networks.

The use of a mobile handset and its applications is now an integral part of our fellow citizens' daily habits. Portable connected devices are increasingly powerful: in many instances they have replaced users' landline telephones, cameras, computers and even televisions. Today, 5 million videos are watched on YouTube and 67,000 images uploaded to Instagram every minute (see Figure 3).
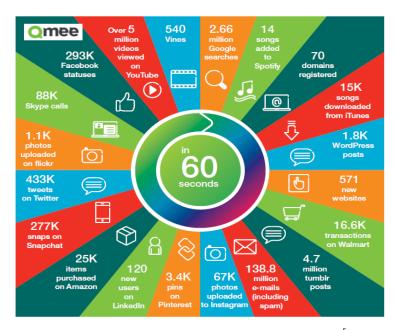
Figure 3. What we do over the network today in one minute[5]

The latest *Mobility Report*[6] from Ericsson indicates that traffic on mobile networks almost doubled in a single year, and that over the next five years it will have increased to 10 times what it is today. New solutions must therefore be found to meet this demand, and to optimise how resources are used.

The increase in the number of applications available, their diversification and the improved quality of mobile networks have all contributed to driving up demand, the emergence of new uses (connected objects, drones, etc.) and new users.

---

[4] On 3 April 1973, Motorola's Martin Cooper made a call from the corner of 56th street in New York City with the first mobile phone.

[5] https://www.ericsson.com/res/docs/2014/5g-what-is-it-for.pdf

[6] https://www.ericsson.com/assets/local/mobility-report/documents/2016/ericsson-mobility-report-november-2016.pdf

5G is at the crossroads of these news uses; it aims to better and simultaneously satisfy this tremendous variety of needs and these new demands, through a unified technology that takes this diversity into account at the design stage.

The advent of 5G could have a significant impact not only in the technical realm, but also on different countries' economic and social development. As indicated in the introduction, 5G targets a very large number of sectors and, through society's digitisation, is expected to contribute to countries' economic growth.

To give an example, according to a report produced by InterDigital Europe, Real Wireless, Tech4i2 and Connect (Trinity College Dublin)[7], this new technology will require a great deal of money and a great deal of work, but will generate €113.1 billion in profits per annum for the European economy by 2025.

1.1 Technical specifications of 5G

The specifications for a new generation of mobile telephony are set primarily by two bodies: ITU (International Telecommunication Union) and 3GPP (3rd Generation Partnership Project).

ITU is the United Nations agency devoted to information and communications technologies. It carries out research and studies through its Working Party 5D, the sub-group responsible for the overall radio system aspects of international mobile telecommunications (IMT). In 2013, this group began working on defining the characteristics of the new IMT standard, IMT-2020 (5G) (cf. 1.3), as it had done back in the early 2000s to define 4G (IMT-Advanced). Further details on the roadmap established for this work can be found in Annex 1, the objective set by ITU-R being to complete its analyses by 2020.

Parallel to the work being done by ITU are the studies being conducted by 3GPP. The 3rd Generation Partnership Project was created in 1998, and its members include seven standardisation bodies, several hundred industry players, associations and public organisations. It is responsible for developing and maintaining technical specifications for mobile telephony standards[8]. When a new standard is being defined by ITU, 3GPP works on the technical solutions that make it possible to achieve the objectives set by ITU.

Although 5G is one of the most widely debated topics inside international and European bodies today, no standard has yet been defined by 3GPP. The draft of Release 15, the first 5G standard from 3GPP, is still in the works: definition of the new architecture began in December 2016 and work on the New Radio (NR) interface is set to begin in March 2017. A first Release of the standard should be validated in September 2018 to meet the more urgent demands; a second Release (3GPP Release 16) will then be published in March 2020. All of these elements are addressed in more detail in Annex 1.

It should also be said that the transition from one generation to the next takes place gradually. LTE will continue to evolve alongside NR, and these two standards will likely be very complementary initially. In particular, for pioneer 5G rollouts LTE shall probably remain the master of the network and control the NR antennae. Moreover, some of the objectives set for 5G could be achieved thanks to functionalities or technologies introduced in 3GPP Releases 13, 14 and 15 that will not be proper to 5G, but rather evolutions of 4G (which some refer to as 4.9G or LTE Advanced Pro).

---

[7] http://ir.interdigital.com/file/Index?KeyFile=36051369

[8] GSM for 2G, UMTS for 3G and LTE for 4G
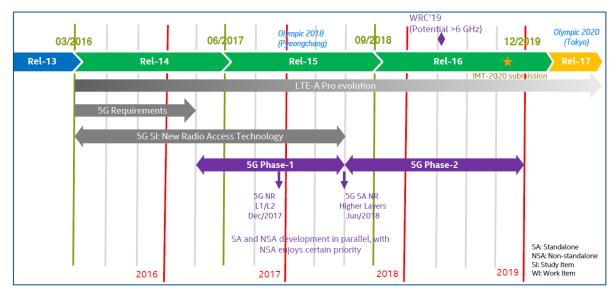
Figure 4. 3GPP timeline for 5G[9]

This same momentum occurred when making the transition from 3G (Release 4) to 4G (Release 10, first Release of IMT-Advanced): the first LTE specifications (Release*s* 8 & 9) do not achieve all of the ITU targets for 4G, and reprised many of the characteristics of the most advanced form of 3G at the time.

The synergies between the successive generations, the absence in most instances of a great initial jump in performances between generations, and the race between competitors to outperform one another and be the most innovative, can often lead operators and equipment suppliers to give a "commercial" name to each generation of mobile telephony. For instance, in the United States AT&T had called its HSPA network 4G, while in France it is a 3G+ network.

So in all likelihood the first 5G networks deployed on a large scale will be 4.9G systems, using carrier aggregation, massive MIMO (Multiple-Input Multiple-Output) and Network Function Virtualization, or NFV (cf. trials conducted in in France, detailed in Section 1.5.3). These technologies, which will be explored in greater detail further on, represent more of an evolution of the fourth generation than an actual transition to 5G, which will occur when disruptive technologies such as NR carriers in millimetre bands, non orthogonal multiple access (NOMA*)* and mobile edge computing (MEC) can be put into place.

In addition to this commercial race, equipment suppliers – often backed by their respective government at home, notably in the United States, Japan and South Korea – want to get a head start in defining and testing 5G technologies, in the hope of establishing themselves as the technological leaders in international standardisation bodies. Here, the precocity of the first rollouts announced in South Korea (5G deployment for the Winter Olympics in 2018 in Pyeongchang) and Japan (5G deployment for the Summer Olympics in Tokyo) would justify the supposition that only a small portion of 5G technologies will be used, and that these rollouts will be based more on 4.9G or pre-5G technologies.

---

[9] http://gsacom.com/paper/intel-5g-technology-mm-wave-frequencies/

1.2   5G use cases

Three main use cases (defined by ITU, under IMT-2020), with their respective – and potentially mutually incompatible – demands are in the process of taking shape, and will make it possible to meet the sector-specific needs referred to in the introduction.

1. **mMTC – Massive Machine Type Communications**: A very large number of connected devices with disparate quality of service requirements. The objective of this category is to provide a response to the exponential increase in the density of connected objects;

2. **eMBB – Enhanced Mobile Broadband**: ultra high-speed connection indoors and outdoors, with uniform quality of service, even on the edges of a cell;

3. **uRLLC – Ultra-reliable and Low Latency Communications**: this use case has stringent requirements for capabilities such as latency and packet-loss, to ensure increased reactivity.

(Source: ETRI graphic, from ITU-R IMT 2020 requirements)
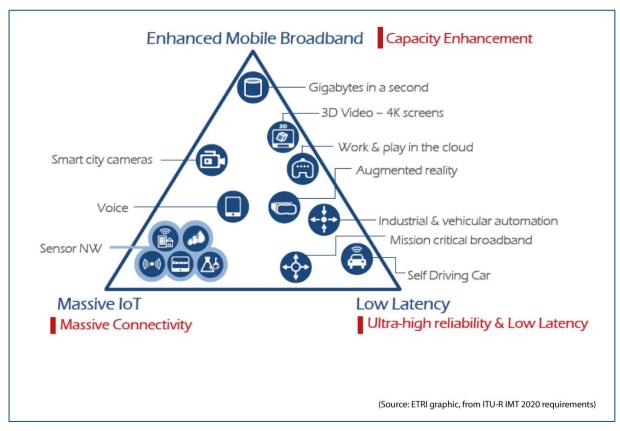
Figure 5. 5G use cases[10]

The first group (mMTC) primarily encompasses all Internet of Things related uses. These services require broad coverage, lower energy consumption and relatively slow transmission speeds. What 5G will deliver compared to existing technologies is the ability to connect objects that are spread out in a very dense fashion across a given area.

Enhanced mobile broadband (eMBB) concerns all of the applications and services that require increasingly fast connections, for instance to watch ultra high-definition (8K) videos or to stream virtual or augmented reality applications wirelessly.

---

[10] http://5g.ieee.org/standards

Ultra-reliable and Low Latency Communications (uRLLC) include all of the applications that require an extreme reactivity and very strong message transmission guarantees. The sectors where these requirements are particularly prevalent are transportation (reaction time when an accident occurs, for instance) and medicine (telesurgery), and when digitising manufacturing processes in general.

To implement these three use case, ITU established eight key performance indicators[11] (KPI) to specify, quantify and measure the characteristics of IMT-2020 (5G) systems:

- Peak data rate (Gbit/s);
- User experienced data rate (Mbit/s);
- Spectrum efficiency (bit/Hz);
- Device mobility (km/h);
- Latency (ms);
- Connection density (number of connected/accessible objects per km²);
- Network's energy efficiency;
- Area traffic capacity (Mbit/s/m²).

---

[11] Added to these eight classes, five new indicators were defined and are currently being examined: reliability, mobility interruption time, bandwidth, maximum spectrum efficiency, 5th percentile spectrum efficiency.

1.3 Jump in performance compared to 4G networks

As mentioned in Paragraph 1.1, the deployment of 5G is likely to take place in two stages:

1. The first 5G rollouts will deliver better performances, but as a continuation of what is being provided at the time by 4G systems which will continue to evolve as well (4.5G, 4.9G);

2. Performances will continue to improve with the gradual introduction of disruptive technologies, such as the use of millimetre wave frequencies.

This gradual rollout is very similar to the way in which 4G was introduced, as the performances obtained with pioneer deployments were relatively similar to those supplied by existing 3G networks.

ITU defines the 4th generation of technologies, called IMT-Advanced, by the values for the eight KPI listed above, as indicated in the following chart. The 5th generation, called IMT-2020, is represented as well. Here, it should be noted that the official ITU definition of 4G in fact corresponds to what commercial rollouts supplied later on, under the name 4G+ or LTE-Advanced.
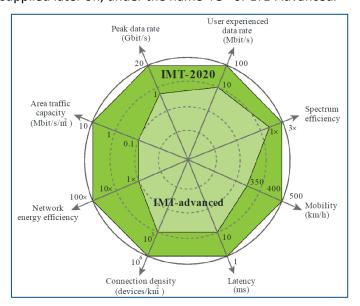
Figure 6. Comparison between 4G and 5G with respect to the eight key performance indicators[12]

According to these objectives, 5G must be able to provide a user experienced data rate and a peak data rate that is respectively 10 and 20 times higher than what is currently available. The maximum connection density will be multiplied by 10 and latency divided by at least 10 (the target point-to-point latency is 1 ms, compared to 30 to 40 ms today).

---

[12] Recommendation ITU-R M.2083-0 (09/2015), https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf
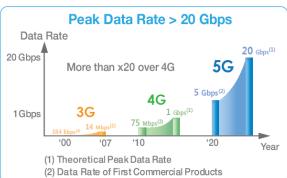
Figure 7. Representation of the "average user experience data rate" KPI for 4G and 5G and "Peak data rate" KPI for 3G, 4G and 5G[13]

As the successor to 4G, the objective for 5G will also be to provide:

- An extremely reliable network, with more consistent performances regardless of the user's position with respect to the base station;
- A stable connection, even when travelling (at speeds of up to 500 km/h);
- Greater network energy efficiency (with batteries that consume up to 100 times less power).

The following table summarises the target performances for 5G and those currently available with 4G:

Performances/Generation	4G	5G
1. Peak data rate (Gbit/s)	1	20
2. User experience data rate (Mbit/s)	10	100
3. Spectrum efficiency	1x	3x
4. Speed (km/h)	350	500
5. Latency (ms)	10	1
6. Connection density (number of objects/km²)	10^5	10^6
7. Network energy efficiency	1x	100x
8. Area traffic capacity (Mbit/s/m²)	0.1	10

Table 1. Comparisons between 4G and 5G performances

---

[13] http://www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-0.pdf

1.4　Network slicing and software-defined networks

It is vital to understand that the set of indicators listed in Section 1.3 determine a set of peak performances for 5G. It will not, however, be possible to achieve all these peak values simultaneously: not every requirement or use case is compatible, so a trade-off will need to be made when defining categories of use that each have their own performance envelope, notably for the use cases described in Section 1.2 (mMTC, eMBB and uRLLC). This is the principle of network slicing: each slice has its own set of KPI, which is a trade-off tied to the target use. On a 5G system, the network properties will need to adapt to the chosen environment.

The following diagram positions the three main use cases listed in Section 1.2. with respect to the eight key performance indicators listed above.
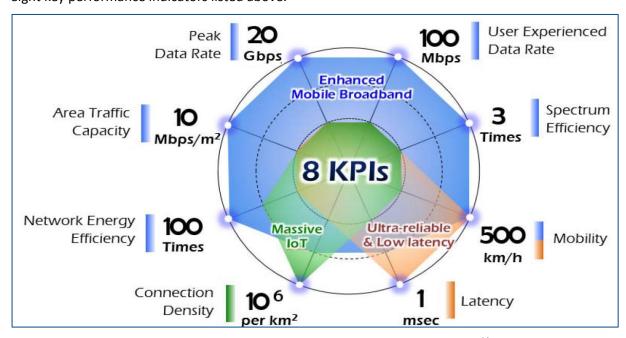
Figure 8. Key performance indicators for the three 5G use cases[14]

So for applications that require enhanced mobile broadband (eMBB), such as 4K, 8K or 3D video or virtual reality, a certain number of performance indicators, such as spectrum efficiency, peak data rate and area traffic capacity, can be reached only at the expense of others, such as latency or connection density.

On the flipside, when a massive simultaneous connection of connected objects (mMTC) needs to be managed, the network will concentrate its resources and use the technologies required to achieve this task, but will not be able, for instance, to use spectrum as efficiently or to guarantee low latency.

Lastly, when ultra-reliable and low latency communications (uRLLC) are required, the number of simultaneous connections, data rates and spectrum efficiency may be reduced.

This flexibility, or ability to adapt, that network slicing brings can only be achieved thanks to the softwarisation and virtualisation of a sizeable number of network components (cf. 1.5.2) – a process referred to as Software-Defined Networking (SDN) and Network Function Virtualisation (NFV). Behind these acronyms is a common idea, namely to use as many generic and reconfigurable components as possible, rather than bespoke ones that are permanently dedicated to very specific

---

[14] https://5g-ppp.eu/wp-content/uploads/2016/11/06_10-Nov_Session-3_Lee-JunHwan.pdf

tasks. This evolution towards software-based systems has been in the works for several years, but is now becoming possible thanks to improved performances from all of these reconfigurable components, including those that are the closest to the elementary tasks of wireless communications (detection, baseband coding, bitstream management, frequency handover, signal processing, etc.).

1.5 Technological building blocks to achieve the objectives

1.5.1 Air interface

Several, sometimes competing radio access technologies are currently being examined. Some have already been pre-implemented by equipment manufactures and can be used in trials, notably massive MIMO and NFV. Others, such as NOMA modulation and mobile edge computing (MEC), will no doubt take longer before they are ready to be used. In any event, a consensus – which could be painful for certain suppliers whose investments will be lost – will need to be found when defining 5G standards, to ensure the systems' interoperability.

The technologies currently being examined are the following:

- **Millimetre wave frequencies**: the use of millimetre wave frequencies constitutes one of the disruptive 5G technologies. The term refers to the frequencies above 6 GHz which have never been considered for mobile fronthaul network rollouts, for reasons of technological maturity and propagation quality. To meet the demand for ever increasing data rates and traffic volumes, new bands with very wide channels (over 100 MHz per user) will need to be employed: millimetre wave frequencies could provide this spectrum resource, and in certain cases their use would make it possible to achieve the data rates listed in Table 1. Comparisons between 4G and 5G). In exchange, to be able to use these frequencies all of the required, miniaturised low-cost technologies will need to be developed, and ensure a level of energy consumption that is compatible with portable devices (amplifiers, coders, signal processing, antennae, etc.). In particular, because of millimetre waves' poor propagation quality, each cell will have limited coverage and so require the use of beamforming (described below) to better focus the power transmitted by the antenna.

- **Massive MIMO** (Multiple Inputs – Multiple Outputs): this technology involves the use of a large number of smart micro-antennae, located on the same panel (between eight and 128 today, but the number will increase with the use of frequencies above 6 GHz). The appeal of using massive MIMO is twofold: first, the technology makes it possible to increase data rates, thanks to spatiotemporal multiplexing; second, it makes it possible to focus energy on a device to improve its link budget, thanks to beamforming.
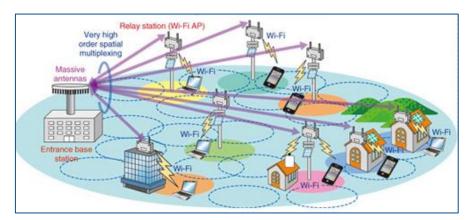
Figure 9. Example of the use of a beamforming antenna to connect Wi-Fi access points[15]

-   **Full Duplex**: in classic systems, transmission and reception takes place either on different frequency bands, i.e. frequency division duplexing (FDD) used on all mobile network bands in France, or at different times: time division duplexing (TDD), the top contender for LTE wireless local loop networks in France. The full duplex is intended to enable the simultaneous transmission and reception of data, on the same frequencies, at the same time and in the same location.[16]
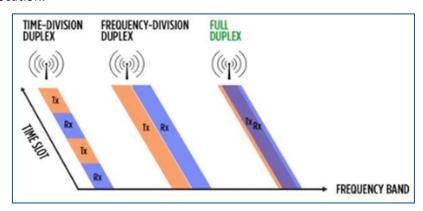
Figure 10. Illustration of full-duplex, compared to FDD and TDD[17]

-   **NOMA Multiplexing** (Non Orthogonal Multiple Access): LTE uses what is referred to as orthogonal multiplexing, with each device using a portion of the resource blocks in a unique fashion at any given time. For 5G to provide improved spectrum efficiency compared to 4G, the plan is to use non-orthogonal multiplexing methods, whereby several users can use the same frequencies at the same time. A distinction can be made between several users by assigning different codes to each user – referred to as SCMA or sparse code multiple access – a combination of 3G's code division multiple access (CDMA) and 4G's orthogonal frequency

---

[15] https://www.slideshare.net/100001290086432/massive-mimo

[16] The basic operating principle is the following: an antenna sends a signal at the same time as it is receiving signals coming from devices in the cell. However, the signal received by the antenna is a combination of the signal sent by itself and signals coming from the devices (everyone is "talking" at once). As it knows which signal it itself has sent, the antenna can subtract it during digital processing from the ones it received. So only the signals received from the devices remain.

[17] http://compeng.columbia.edu/biggest-component-full-duplex-wi-fi-radio-antenna

division multiple access (OFDMA) or by playing on the difference in users' signal to noise ratios (power domain NOMA[18], illustrated below). These are the two methods chosen for NR.
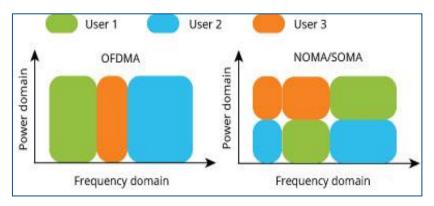
Figure 11. Illustration de multiplexing *power domain* NOMA[19]

- **QAM256**: as with many modern communication systems, 4G uses quadrature amplitude modulation (QAM). In 4G this modulation can achieve QAM64, which means that six bits of information are being transmitted (2^6 = 64) at any given time. On wireless systems, the main limitation on QAM order is the signal to noise ratio: when a large amount of information is sent all at once, its transmission will be very sensitive to disruptions (a bit like trying to talk in a noisy environment: it is easy to understand "yes" or "no" but harder to understand more complex sentences). Thanks to an improved link budget, via antenna or signal processing technologies, 5G modulation could reach QAM256, i.e. eight bits of information being transmitted at any given time, which translates into a 33% increase in maximum capacity under ideal conditions. This improved modulation will also be deployed on advanced 4G systems.

- **IoT waveforms**: new waveforms are being explored for the future deployment of 5G IoT in mobile bands. But although mass market IoT is one of the main challenges put forth for 5G, no concrete results have yet been made public. Operators[20] are starting to deploy new standards (EC-GSM or Extended Coverage GSM, LTE-(e)MTC or enhancements for Machine-Type Communications, NB-IoT or NarrowBand IoT) which were defined by 3GPP in Release 13 but, as they are based on 2G and 4G, they do not deliver the performance levels, notably in terms of autonomy, coverage and density, that are compatible with the targets set for future 5G networks.

1.5.2   Network architecture

As with air interfaces, new network architectures are also being explored:

- **Software-defined networking (SDN) and network functions virtualisation (NFV)**: these two functionalities extend beyond the scope of 5G networks per se. They are part of an overall

---

[18] The basic operating principle is the following: either User 1 (U1) with a good signal to noise ratio and User 2 (U2) with a less good signal to noise ratio. The antenna sends a high power Signal 2 to U2, and superimposes a weaker Signal 1 aimed at U1. U2 will only see S2 as S1 is drowned out by the noise. U1 will decrypt S2 then delete it from the signal received, to create a higher quality S1, thanks to a better signal to noise ratio.

[19]   https://www.anritsu.com/en-AU/test-measurement/technologies/5g-everything-connected/5g-everything-connected-detail

[20] http://www.vodafone.com/content/index/what/technology-blog/nbiot-commercial-launch-spain.html

process of network upgrades taking place today and already available with 4G technology (4.9G). The nevertheless remain a key enabler of 5G:

o SDN (Software-Defined Network) is designed to disassociate the network's control plane from its data plane, these two planes traditionally being linked and distributed in a set fashion in the network (see diagram below). Controlling the network, a task previously assigned to specialised and unscalable hardware components, is centralised in the form of software on more powerful servers and, in theory, free of equipment manufacturer specifications. This enables the deployment of high value-added services (load balancing, smart routing, dynamic configuration, etc.) in disparate environments.
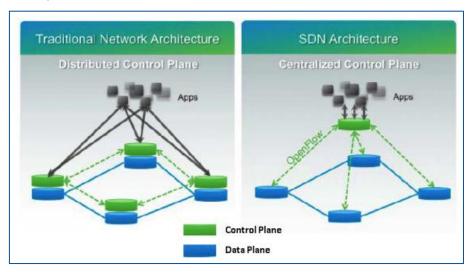
Figure 12. Centralisation of the control plane in a software-defined network[21]

o NFV, which builds upon SDN, is used to virtualise, in other words to replace hardware designed specially to perform certain key network functions (firewall, network core, interfaces between different systems…) with software on a server, to accelerate rollouts and enable rapid changes and upgrades.

- **CloudRAN**: this functionality, also know as centralized-RAN, requires a very different network architecture to what we find today. It is an evolution of SDN: the base stations' signal processing units, currently installed at the base station level, are moved to the cloud and centralised. They communicate with the network radio heads, located closer to the antenna, over an optical fibre network (Radio over fibre technology). This centralisation makes it possible to obtain a complete overview of all of the stations deployed and to coordinate signal processing and manage interference between cells and devices

---

[21] http://www-igm.univ-mlv.fr/~dr/XPOSE2014/software-defined_networking/sdn.html
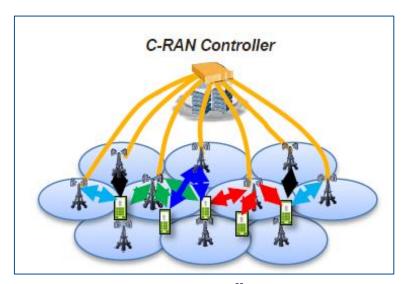
Figure 13. Illustration of CloudRAN[22] network architecture

- **Optimised content delivery, using a mobile content delivery network (mobile CDN)**: corresponds to a set of servers working together in a transparent fashion to optimise the delivery of content to end users over wireless (mobile or Wi-Fi) networks, with high availability and performance. With 5G, the objective for these CDN is to cache content close to users, notably thanks to predictive algorithms, to offload traffic from the networks and decrease latency.

- **MEC (mobile edge computing)**: MEC is an evolution of mobile CDN whose purpose, in addition to bringing data closer to devices, is to provide devices with an accessible computing power with very low latency, within a very specific area for demanding applications. This technology makes it possible to locate a portion of the network's intelligence (managing local critical applications and performance analysis) at the base station level. The "antennae" will be capable of analysing a certain number of data, and so to make decisions very quickly.

- **Device-to-device**: D2D is a direct form of communication between two nearby devices, which does not require the data to travel over the cellular network. Device-to-device communication is not new, as technologies such as Bluetooth and Wi-Fi direct already enable it. But a new mesh networking technology will be introduced with 4.9G and later 5G network rollouts: LTE-direct. Far more energy-efficient than its predecessors, this technology will have a range of up to 500 metres and geolocation capabilities to enable communications. This technology will be very useful for low latency V2V (vehicle-to-vehicle) or V2X (vehicle-to-everything) communications and for certain public-security related uses.

1.5.3 5G: a multi-technology generation cohabitating with existing networks

As stated earlier, 5G is not meant to replace 4G overnight. In practice, the devices will undoubtedly be multi-modal: still connected to the 4G network, which will provide extended coverage for pioneer rollouts, then transitioning to 5G networks when they become available.

While 4G and 5G frequencies will probably be initially segregated, in all likelihood devices will rapidly become capable of aggregating 4G and 5G carriers and, further down the road, 4G carriers will be encapsulated in 5G ones.

---

[22] https://5g-ppp.eu/wp-content/uploads/2016/11/04_10-Nov_Session-3_Takaharu-Nakamura.pdf

4G is still being deployed, and its technological evolution and certain building blocks will be used by both the first 5G networks and advanced 4G networks. To wit, the latest trials conducted in France are allowing 4G networks to perform better thanks to the use of pre-5G technologies:

- Bouygues Telecom, in partnership with Huawei, managed to achieve a peak data rate of 1 Gbps thanks to the simultaneous use of four-carrier aggregation (800 MHz, 1800 MHz, 2100 MHz, 2600 MHz) and more powerful modulation (256 QAM)[23];
- In the coming weeks, Orange will be launching a massive MIMO (16x16) trial with Nokia[24].

In addition to the strong integration between 4G and 5G, the new generation will no doubt also continue convergence efforts between frequency bands governed by exclusive licences – i.e. bands that are allocated exclusively to an operator, such as mobile operators – and unlicensed frequency bands, governed by a system of general authorisation (e.g. Wi-Fi bands), which already began in 4G with LTE-LAA (Long Term Evolution – License Assisted Access) and LTE-LWA (Long Term Evolution – Wi-Fi Link Aggregation).

LTE-LAA is characterised by the aggregation of one or several LTE carriers, used in licensed bands, with other LTE carriers employing unlicensed 5 GHz Wi-Fi bands. To guarantee cohabitation with very widespread Wi-Fi networks whose deployment pattern is unpredictable, it uses LBT (Listen Before Talk) technology to listen to the radio channel before transmitting, in order to determine whether or not a frequency is available.

LTE-LWA consists of an aggregation of LTE carriers in licensed bands with Wi-Fi traffic. To achieve this, the LTE cell and Wi-Fi access point need to be connected: the Wi-Fi traffic is sent to the cell which sends back the entire aggregated link on the 4G core network. This technology is particularly well suited to indoor environments with small cells, a system which, as detailed in paragraph 3.3 below, will lend itself well to 5G deployments.

---

[23] https://www.corporate.bouyguestelecom.fr/wp-content/uploads/2016/05/Communiqu%C3%A9-Bouygues-Telecom-et-Huawei-Technologies-r%C3%A9alisent-un-test-de-d%C3%A9bit-4G-%C3%A0-plus-de-1-Gbps-pour-la-premi%C3%A8re-fois-en-Europe-de-lOuest.pdf

[24] https://networks.orange.fr/actualites/actualites-des-networks/orange-et-nokia-partners-pour-developper-les-futurs-services-5g

2   5G development initiatives

With a view to keeping the many promises being made for 5G, a number of initiatives are currently underway around the globe to promote its development. Below, we detail the foremost initiatives put forth by the players with which Arcep met as part of its investigation.

2.1   Government initiatives

The prospect of significant socio-economic repercussions generated by 5G (possible revenue of close to $225 billion a year[25] by 2025) combined with many countries' desire to establish themselves as technological leaders and make their companies more competitive, have propelled a multitude of government initiatives around the globe, aimed at encouraging the mobile ecosystem to begin the work and make the investments required to drive the rapid construction of the first 5G networks.

A selection of the largest initiative is detailed below.

2.1.1   In Europe

5G-PPP

The 5G Public Private Partnership (5G-PPP) is dedicated to 5G research and development, created on the initiative of the European Commission in 2013, with a budget de €700 million in public funding. The main objectives set by 5G-PPP are:

- Create stronger ties between the economic players and academic bodies devoted to the telecommunications sector over R&D projects, along the entire value chain;
- Reduce technological dependence on the United States and Asia while sustaining a strong global market;
- Regain technological leadership, notably in disruptive technologies, by promoting standards in international bodies;
- Allow innovative business models to emerge;
- Facilitate large-scale experimentation.

The results of this work will help clarify the 5G action plan (see below), and fuel the standardisation work that is currently underway.

To achieve its ambitions, 5G-PPP has initiated three stages of work, financed by the European Union, whose roadmap is in sync with the main international initiatives (3GPP and ITU, cf. Figure 14):

- The first stage which is currently underway, will last until mid-2017;
- A second stage focused on systems optimisations, from the end of 2017 to mid-2019;
- And finally a full-size trial stage from 2019 to 2020.

Deploying 5G by 2020 will require Europe to develop leading edge technologies, globally approved standards and especially to achieve consensus over the most suitable frequency bands. This funding and these projects – involving a great many researchers from more than a hundred companies and the finest R&D centres in Europe – are thus vital.

---

[25] https://www.abiresearch.com/press/abi-research-projects-5g-worldwide-service-revenue
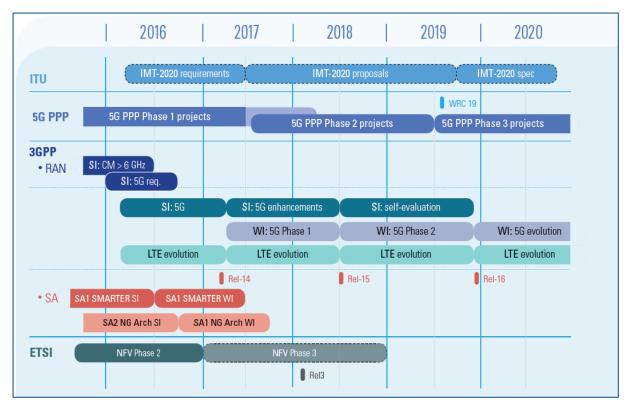
Figure 14. 5G PPP vs. 3GPP and ITU roadmaps[26]

5G action plan

As a complementary measure, on 14 September 2016 the European Commission launched its *5G for Europe Action Plan* to bolster investments in 5G infrastructure and service rollout efforts in the Digital Single Market between now and 2020. This action plan sets out a clear roadmap for public and private 5G investments inside the EU.

The Commission has proposed the following measures to achieve this plan:

- Align roadmaps and priorities for a coordinated 5G deployment across all EU Member states, targeting early network introduction by 2018, and moving towards commercial large scale introduction by the end of 2020 at the latest.
- Make provisional spectrum bands available for 5G ahead of the 2019 World Radio Communication Conference (WRC-19), to be complemented by additional bands as quickly as possible, and work towards a recommended approach for the authorisation of the specific 5G spectrum bands above 6GHz.
- Promote early deployment in major urban areas and along major transport paths.
- Promote pan-European multi-stakeholder trials as catalysts to turn technological innovation into full business solutions.
- Facilitate the implementation of an industry-led venture fund in support of 5G-based innovation.
- Unite leading actors in working towards the promotion of global standards.

---

[26]5G Empowering vertical industries. White Paper, 2016, https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf

The European Commission has given every EU country a certain number of ambitious, numerical targets. One core objective for 5G is thus to have at least one major city in every European country outfitted with this new generation mobile system by 2020, and coverage of every city, motorway and high-speed railway lines by 2025. This comes in response to announcements from South Korea and Japan which are both promising large-scale 5G demonstrations, respectively, at the Winter Olympics in Pyeongchang in 2018 and the Summer Olympics in Tokyo in 2020.

2.1.2 Worldwide (examples in the US, South Korea, Japan and China)

The race is already underway between countries to be the first to begin large-scale 5G trials, and later to introduce commercially available services, so much so that they are willing to employ pre-standard technical specifications.

Several countries are therefore working on standards and seeking to reach a consensus with one another to achieve international backing for the technical specifications that will satisfy their needs, and earn them the best possible return on the investments made thus far. This in turn is creating a certain turmoil, for instance, over the frequency bands that would be the best candidates for 5G.

The United States[27]

5G is seen as an unprecedented opportunity for economic growth, with a tremendous impact on education, job, transportation, etc. According to the US federal regulator, the FCC (Federal Communications Commission), the following three elements in particular need to be the prime focus of attention: spectrum, infrastructure and the backhaul network:

- **Spectrum**: in July 2016, the FCC voted to free up and to open up nearly 11 GHz of high frequency spectrum to be used for fixed and mobile broadband applications: 3.85 GHz of licensed spectrum in the 27.5 – 28.35 GHz and 37 - 40 GHz bands, as well 7 GHz of unlicensed spectrum from the 64 - 71 GHz band. The FCC's stated goal was to provide the assurance and clarity for investments in the telecoms sector. Verizon welcomed this decision, and plans on achieving the first rollouts in 2017, with trials already underway in several cities around the country (cf. 3.3).

- **Infrastructures**: The FCC believes that 5G must be underpinned by a robust infrastructure network capable of handling already heavy traffic that could potentially increase exponentially in the coming years. It will interact in a hybrid fashion between traditional towers with macro cells and small cell deployments, as well as distributed antenna systems. With an eye on the deployment of new towers, the FCC resolved to reduce, or eliminate in certain cases, the regulatory restrictions on installing facilities and antennae that will have very little impact "on historic properties" (tower re-use, small antennae, indoor installations, etc.)[28].

- **Backhaul network**: Because 5G systems will require very high-power backhaul networks, the FCC is in the process of reforming and updating the regulations governing data services markets, including wireless backhaul solutions. The FCC is seeking to protect this market while also creating competition, so that competitive and high quality backhaul network solutions can emerge.

---

[27] Forging Our 5G Future – Federal Communications Commission: https://www.fcc.gov/5G

[28] https://apps.fcc.gov/edocs_public/attachmatch/DA-16-900A1_Rcd.pdf

Japan

Japan wants to demonstrate its 5G leadership by deploying the first commercial 5G network that complies with international technical specifications, in time for the Summer Olympic Games in Tokyo in 2020.

According to the *Radio Policies Towards 2020*[29] report, published in June 2016 by Japan's Ministry of Internal affairs and Communications (MIC), the 3600 - 4200 MHz, 4400 - 4900 MHz and 27.5 – 29.5 GHz bands were selected to be the 5G candidate bands on a national scale. Although other bands are also being investigated, with a view to WRC19, 5G rollouts are already planned in the 3600 - 4100 MHz, 4405 - 4895 MHz and 27.5 - 28.28 GHz bands as early as 2017 in Tokyo, and will be built out in 2018 and 2019.

China

Like Japan, China too wants to prove its leadership with the first commercial rollouts in 2020. The first trials will be conducted that year in the 3400 – 3600 MHz band. The 3300 - 3400 MHz, 4400 - 4500 MHz and 4800 - 4990 MHz bands are also under consideration and currently being investigated. For higher speeds, the country plans on using spectrum around the 25 GHz and 40 GHz frequencies.

South Korea

South Korea, meanwhile, has set its sights on a pre-commercial 5G service that would be ready in time for the 2018 Winter Olympics in Pyeongchang. Several trials are already underway to prepare for full scale demonstrations in several South Korean cities, including Pyeongchang and Seoul. The country's three national mobile operators are hoping for spectrum in the 26.5 – 29.5 GHz bands.

South Korea's largest mobile operator, SK Telecom, has announced a plan for conducting interoperability tests with Qualcomm and Ericsson, as well as outdoor trials of the new air interface based on the 3GPP NR standards that are being developed today. These trials and experiments will be carried out in the second half of 2017.

If the goal of these trials is to help accelerate the specification of the new NR air interface, which is part of the work being done on 3GPP Release 15, it has not yet been specified which frequency bands will be used for these trials.

2.2 A host of private initiatives – a few examples

5G Open Trial Specification Alliance

In early 2016, South Korean operators KT and SK Telecom, Japanese carrier NTT DoCoMo and American carrier Verizon formed the 5G Open Trial Specification Alliance to carry out collaborative 5G trials.

Due to be conducted between 2016 and 2018, the aim for these trials is to provide a common platform for operators, for exchanging results and shared assessments of the different 5G network components and elements. One of the operators' objectives is to help accelerate the definition of standards and specs, and to enable economies of scale. The findings of these evaluations will provide input for 3GPP discussions, adding simulations to the experimental data produced by the partnership.

---

[29] http://www.gsma.com/spectrum/wp-content/uploads/2016/08/MIC_Spectrum-for-5G-MIC-Kuniko-OGAWA.pdf

The alliance's founders hope to attract a number of industrial partners to the platform, including other carriers, equipment suppliers, chipset makers, etc. Several of the founding operators' equipment suppliers are already involved in the scheduled trials.

The trials will cover multiple frequency bands, both above and below 6 GHz.

Verizon 5G Technical Forum

The Verizon 5G Technology Forum (V5GTF) was created in late 2015 by Verizon, in cooperation with its partners Cisco, Ericsson, Intel, LG, Nokia and Qualcomm. The goal of the collaboration is to provide a platform for testing radio interface specifications in the 28 and 39 GHz bands.

The first fruit of this collaboration, in July 2016 Verizon announced the completion of the first specifications of their 5G radio interface. These specs should enable the different parties involved to develop interoperable solutions and so to help drive forward the definition pre-standards.

Even though they run the risk of developing solutions that are not compatible with 3GPP or ITU standards, Verizon believes it has a solid enough grasp of the overall concepts being discussed in those bodies. The tests carried out in several US cities were thus able to validate substantially greater capacities than 4G.

Orange/Ericsson

Orange and Ericsson have been working together since October 2016 on developing 5G use cases and services, as well as demonstrations. The purpose is to develop technological building blocks, conduct trials and pilot projects for a range of use cases, including multi gigabits/s wireless internet access in suburban and rural environments, the Internet of Things and connected cars.

Their collaboration made it possible to achieve 15 Gb/s in a laboratory environment, notably thanks to the use of massive MIMO and beamtracking using centimetre waves. The collaboration will also focus on the transition from 4G to 5G solutions, notably in terms of reducing costs and improving energy efficiency, and on the use of SDN and NFV technologies.

Since January 2017, the partnership has included PSA (Peugeot) to carry out trials on connected cars.

5G-ConnectedMobility

5G-ConnectedMobility is a consortium formed by Ericsson, BMW, Deutsche Bahn, Germany's three mobile operators, Deutsche Telekom, Telefonica Deutschland and Vodafone, the 5G TU Dresden laboratory, the BAST research institute and BnetzA, with the goal of galvanising and accelerating 5G R&D in Germany.

5G-ConnectedMobility thus plans on supplying a digital motorway infrastructure and real application environment to be able to test V2V (vehicle-to-vehicle) technologies as well as solutions for digitising railway infrastructure.

To this end, 5G-ConnectedMobility operates with the help of an independent network infrastructure that is not connected to any commercial network. The Ericsson mobile network devoted to this project makes it possible to run live tests: Ericsson thus obtained permission from regulator BNetzA to employ spectrum in the 700 MHz band in the Nuremberg area.

3   The challenges of 5G

In this part we explore the different challenges that lie ahead for 5G, which emerged from the interviews that Arcep conducted over the past several months.

3.1   New business models focused on vertical markets

3G and especially 4G technologies were designed primarily for the superfast mobile internet. 5G continues on in this direction, but also wants to target what are known as vertical markets, which encompass several segments, including:

- Connected vehicles, not only to deliver entertainment and information to passengers, but also to guarantee safety via communications both between vehicles and between vehicles and infrastructure;
- Factories of the future;
- Smart cities with requirements in the areas of public transportation (similar to the needs of connected vehicles), the environment, managing buildings and energy consumption;
- Medicine, healthcare and robot-assisted telesurgery;
- Smart grid flow monitoring and management (electricity, gas, water, etc.).

This section will focus in particular on the connected car and factories of the future segments. Because of their current and future macro-economic context, along with the plurality and effervescence of the pioneer work being done in these areas, these two segments constitute the main avenues of 5G development in vertical markets. Smart cities and smart grids have already begun to soar through existing Internet of Things (IoT) technologies.

3.1.1   The automotive sector

The car is an extremely common form of transportation, and safe driving is a fundamental consideration: human error is the number one cause of all transport accidents. The transportation sector wants to use technological innovations to tackle this problem, and to continue to make transportation more efficient, more sustainable and safer.

There are three areas in which technological progress could help improve automotive transportation. 5G could have a role to play in all three, but particularly in the first two:

- Provide in car internet connectivity, to deliver entertainment to passengers;
- Provide access to driver assistance information, to reduce accidents and improve the fluidity of traffic;
- And, finally, the ability to make cars autonomous, thanks to artificial intelligence algorithms.

The first area is just an extension of the developments occurring today around 4G. The aim is to give passengers access to their messaging services, to the internet, to multimedia content, online gaming, etc. The increased connection speeds promised by 5G will improve the use of all of these services.

The second area aims to make cars more intelligent by using information that was not previously available to them. This in turn will help improve the safety and efficiency of the networks as well as help drivers make the right decisions, and adapt to driving conditions. Such connected vehicles could have access to information about dangers on the road (slow or stopped cars, traffic jam warnings, indications of where construction is taking place on the roads, weather conditions, emergency braking, emergency service vehicles approaching, etc.) or regarding signage (signalling/signage on-

board vehicles, speed limits on vehicles, failure to heed signage/safety precautions at crossings, request for right-of-way at traffic lights for designated vehicles, green light optimal speed advisory, etc.). Other services, such as information on refuelling or recharging stations, vulnerable road user protection, managing street parking and traffic information, and smart guidance, could also prove useful. It is still not clear whether, to achieve this, vehicles will simply exchange information with each other or if connectivity with infrastructure will be required along the roads to optimise vehicles' behaviour. In both instances, 5G could have a role to play.

Of the many initiatives that are already underway we can begin with an example from France: the SCOOP@F[30] (cooperative intelligent transport systems) project coordinated by the Ministry of the Environment, Energy and the Sea, and which unites local authorities and R&D centres. Launched in February 2014, new partners subsequently joined the initiative, including Orange and Austrian, Spanish and Portuguese partners. As it is a European project, it receives 50% of its financing from the European Commission, and cross-tests are conducted with Austria, Spain and Portugal. SCOOP@F is a pilot rollout project for cooperative intelligent transport systems; it aims to deploy 3,000 vehicles on 2,000 km of roadway spread across five locations: Ile-de-France, the A4 motorway, the Isère, and ring roads in Bordeaux and Brittany. Its main objectives are to improve road safety and the safety of roadway workers, achieve more efficient traffic management, reduce pollution, streamline infrastructure management costs and participate in defining the car of the future.

In addition, in early 2017, mobile carrier Orange, equipment supplier Ericsson and car-maker PSA Group signed a partnership agreement[31], as part of the "*Towards 5G*" initiative, to conduct technical trials relating to 5G. The aim of this alliance is to test the different paths of technological evolution from 4G to 5G to meet the needs of connected cars, notably in terms of intelligent transport systems (ITS), for safer driving and new on-board services.

The third area concerns the emergence of autonomous vehicles. A number of projects are underway in this area. The first step is to outfit vehicles with algorithms that enable it to make decisions quickly based on their environment. This requires a large number of sensors to deliver a full "understanding" of what is happening around the vehicle. Without pre-judging the technologies that will ultimately be employed to achieve this, as with a human driver (see above), the car could take advantage of connections with other vehicles on the road and with an infrastructure, to have access to all of the aforementioned information.

3.1.2   Industry 4.0

Competitiveness does not depend solely on innovation and the evolution of products, but also on modernising businesses and their means of production. According to certain studies carried out in 2015[32], the digital transition in Europe will enable enterprises to increase their revenue by more than 110 billion euros a year over a five-year period.

A great many countries have set out a strategy for modernising their manufacturing infrastructure (l'*industrie du futur* in France[33], Industry 4.0 in Germany…) of which one aspect will involve digitising processes and trade. The European Commission itself introduced measures in 2016 for strengthening

---

[30] http://www.scoop.developpement-durable.gouv.fr/spip.php?page=sommaire

[31]   http://www.orange.com/fr/Press-Room/communiques-2017/Ericsson-Orange-et-le-groupe-PSA-partners-pour-la-voiture-connectee-en-5G

[32] PwC, "Industry 4.0: Opportunities and challenges of the industrial internet" (2015), and Boston Consulting Group, "Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries" (2015)

[33] http://www.economie.gouv.fr/files/files/PDF/industrie-du-futur_dp.pdf

competitiveness in Europe, in which 5G could play a significant role. This included earmarking 500 million euros for the Horizon-2020[34] research programme.

The advent of new technologies (4G, fibre and soon 5G) and new services (the Internet of Things, the cloud, big data) should facilitate businesses' digital transition. 5G in particular is expected to be a very versatile technology, capable of undergirding a very wide array of uses, and could go a long way in furthering companies' transition to digital technologies and solutions.

3.2 Spectrum harmonisation

5G is emerging as a technology that will use both low frequencies (f < 1 GHz), high frequencies (1 GHz < f < 6 GHz) and, for the first time ever in consumer networks, very high frequencies referred to as "millimetre wave" frequencies (f > 6 GHz).

This spectrum diversity is entirely bound up with the promises of 5G: extended coverage (low frequencies), ultra high speeds (very large channels in very high frequency bands), low power consumption. Furthermore, satellite services will also contribute to the development of this new technology, especially in areas that are difficult to cover and to provide backhaul solutions. The satellite industry is thus taking an interest in 5G, and wants to be involved in defining these new generation network.

3.2.1 Millimetre wave frequencies

The "millimetre" band, also referred to as millimetre wave spectrum, aka frequencies above 6 GHz, are essential to enabling 5G to mark a departure from 4G, for the reasons cited in Section 1.5.1.

At the latest World Radiocommunications Conference (WRC-15 in Geneva), a conference under the aegis of ITU whose objective is to change the way frequencies are allocated between users, discussions over the definition of future mobile bands made it possible to focus future 5G studies, for millimetre wave frequencies, on a certain number of bands situated between 24 GHz and 86 GHz (33.25 GHz identified in total): 24.25 - 27.5 GHz, 31.8 - 33.4 GHz, 37 – 43.5 GHz, 45.5 - 50.2 GHz, 50.4 - 52.6 GHz, 66 - 76 GHz, 81 - 86 GHz.

It is important to stress that, even if the above-listed bands have been identified as "5G bands", at this stage there is no way to know whether they can actually be used to deploy this new generation system: only the results of technical studies will make it possible to establish the constraints and rules of compliance, and to validate the feasibility of these hypotheses.

Figure 15. Millimetre wave frequencies identified at WRC-15

Contrary to the conclusions of the Conference, which reflect European recommendations, the United States and certain Asian countries (South Korea, Japan) have decided to perform 5G trials in the 28 GHz band, and equipment suppliers such as Qualcomm and Samsung, have begun manufacturing 28 GHz- band products.

---

[34] https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020

Meanwhile Europe, following the publication of an RSPG (Radio Spectrum Policy Group within which France is represented by ANFR) opinion[35], decided to focus its first studies on the 26 GHz band (pioneer band), then on the 32 GHz and 42 GHz bands. Later, studies will be carried out on introducing 5G in all of the other bands identified by WRC-15.

The rapid choice of the 26 GHz band as the pioneer band was made to enable economies of scale for equipment production, since it is very likely that dual-mode equipment, i.e. compatible with both the 26 GHz and 28 GHz band, will be available for pioneer rollouts[36].

In France, the 26 GHz band is already being employed for a variety of uses: mobile operators' wireless fixed links (4G infrastructure links), fixed satellite service systems and ground stations for space services. Studies will therefore need to be conducted to take these services into account, and define either the cohabitation or migration of applications to other millimetre wave frequencies.

3.2.2　Frequency bands below 6 GHz

a)　The 3.4-3.8 GHz band

It will not be possible for 5G to run entirely on millimetre wave frequencies: the propagation qualities of these bands make it difficult to achieve widespread coverage, particularly in more sparsely populated areas. In addition, these bands are still lacking in technological maturity when it comes to delivering consumer market communication services. As a result, a "core" band below 6 GHz needs to be identified: one that provides sufficiently large channels to enable future 5G operators to provide innovative services and a higher quality of service than with 4G.

The 3400 - 3800 MHz band appears to be a good candidate.

First, it has already been harmonised for ultrafast mobile services inside the European Union. Initially the 3400 - 3600 MHz band and later the 3600 - 3800 MHz band were identified as the "IMT" bands (for high-speed mobile). Second, they have a great deal of available spectrum (up to 400 MHz). Lastly, technological advancements (antenna and signal processing) make these frequencies compatible with their use to establish macro cells, and not only microcells. To give an example, it emerged from the interviews conducted by Arcep that coverage with this band could be similar to coverage with the 2.6 GHz band, the core 4G frequency band.

This analysis has been confirmed by the European Commission (in its 5G mandate to RSCOM[37]) and by RSPG[38] which consider the 3.4 - 3.8 GHz band as the only credible 5G band for deployments taking place before the end of 2020.

In France, the band is assigned to Arcep, in a priority fashion, for the 3400 - 3600 MHz frequencies (and in a non-priority fashion to the Ministry of the Interior and the Ministry of Defence), and exclusively for the 3600 - 3800 MHz frequency bands. It is used for wireless local loop and satellite applications.

The compatibility of future uses with current applications is being examined by the Electromagnetic Compatibility Advisory Committee, which is chaired by the National Frequency Agency and of which

---

[35] http://rspg-spectrum.eu/wp-content/uploads/2013/05/RPSG16-032-Opinion_5G.pdf

[36] The tuning range is vital here: a system capable of regulating its frequency from 26 to 28 GHz appears entirely technologically achievable within the timeframe for the first 5G millimetre wave deployments. This would have been less true if a switch between 28 and 32 GHz were required.

[37] https://circabc.europa.eu/sd/a/448dc765-51de-4fc8-b6e0-56ed6a1d0bca/RSCOM16-40rev3%205G%20draft_mandate_C EPT.pdf

[38] http://rspg-spectrum.eu/wp-content/uploads/2013/05/RPSG16-032-Opinion_5G.pdf

Arcep is a member: exclusion zones, protecting certain locations tied to long-term use of satellite solutions, need to be set up. Military radiolocation systems operating below 3.4 GHz also need to be protected.

b)   Other bands below 6 GHz

The bands being used today for 2G, 3G and 4G could be used for future 5G deployments.

Refarming 2G, 3G and 4G could be a delicate undertaking, however, because of the duplexing methods they use. Most mobile communications in Europe use FDD (Frequency Division Duplexing)[39] to exchange information. Technical discussions over 5G are nevertheless anticipating that TDD (Time Division Duplexing)[40] will be the chief, if not sole, form of duplexing used for this new generation, notably because it makes it possible to adjust bandwidth to data rates and because it is particularly efficient when beamforming is used.

For 5G networks, using already harmonised mobile bands will thus require in-depth technical studies to define the terms and conditions of use and sharing with existing services. On this point, the ECC (Electronic Communications Committee[41]) decided to assess the potential of certain already harmonised bands, notably the 700 MHz and the L band (1427-1492 MHz).

In France, Arcep allocated the 700 MHz band to mobile operators in late 2015. Even if the country's four operators all obtained frequencies in this band, only Free Mobile, which has no blocks of 800 MHz spectrum, decided to begin its rollouts in the band using LTE (4G) technology.

Arcep is currently analysing the contributions to its public consultation, addressing among other the 3.5 GHz band, that ended recently.

The L band (1427 - 1518 MHz), which was defined to be used exclusively in SDL (Supplemental DownLink) mode, could be considered to meet the constant demand for ever higher data rates and the greater increase in downlink traffic, compared to uplink traffic. In France the band is used by wireless fixed links authorised by Arcep, by the Ministry of Defence for mobile services (excluding aeronautical) and by the Ministry of the Interior. Using it for 4G or 5G would therefore suppose moving the above-mentioned uses over to other bands.

3.3   Increasingly small cells

Today, mobile network rollouts are essentially based on the use of macro base stations: installations that are equipped with high-power antennae deployed to guarantee coverage for a relatively wide area, providing good quality of service. Network configurations are evolving constantly: new radio base stations are installed on a regular basis to increase the networks' capacity, to better meet users' needs in terms of indoor and outdoor coverage, and to improve quality of service.

However, this continually growing demand for capacity already requires operators to increase the density of their networks with smaller and smaller cells.

5G – which will probably bring about a sizeable increase in data traffic, and which will use millimetre wave frequencies whose propagation capabilities are weak – will no doubt require the widespread deployment of low-power base stations (small cells).

---

[39] A technique that consists of sending and receiving data simultaneously, but on two different frequency bands.

[40] A technique that consists of sending and receiving data on the same frequency band but at different times.

[41] The European body that specifies the technical terms and conditions governing the use of frequency bands, which are then set by European Commission decisions.

To satisfy demand and enable the introduction of 5G, estimates indicate at least 10 small cells per macro base station in urban settings[42], where cells are already today relatively tightly meshed to handle traffic density.

Here, the players that met with Arcep raised several questions that will no doubt need to be answered when considering the ubiquitous deployment of small cells.

3.3.1 Taxation

The current regulatory framework stipulates that base stations whose power requires an opinion, an agreement or a statement from the National Frequency Agency (ANFR) are subject to a flat tax on network companies (IFER[43]). The amount of the tax varies according to transmission power, the type of installation and the deployment location. For a deployment in an urban area, the tax stands at €1,607/year/installation for a base station with an effective isotropic radiated power (EIRP) of more than 5W (a COMSIS agreement from ANFR is required to be able to transmit) and €160.70 €/year/installation for an EIRP of between 1W and 5W (declaration to ANFR is required to be able to transmit).

5G small cells will likely use variable transmission powers of between 1W and 25W. In light of rollout density forecasts, hence the number of small cells to install, some of the stakeholders that Arcep met with raised the question of possibly adapting this tax in such a way as to enable massive small cell deployments without generating an equally massive increase in the total tax amount.

This process is already underway, notably with a view to decreasing taxes on base stations in locations that are hard to cover: Act No. 2016-1888 of 28 December 2016 on the modernisation, development and protection of mountain regions thus exonerates mobile base stations built in mountain regions between 1 January 2017 and 31 December 2020 from paying the IFER tax.

3.3.2 Access to elevated and "semi-elevated" locations

To perform their deployments successfully, mobile operators have traditionally needed to install their base stations in elevated locations (towers, rooftops, etc.). This will continue to be true with 5G networks, but will be even more challenging for two main reasons:

1. 5G antennae will probably be larger (in the m^2 range for some) than current 2G, 3G or 4G antennae, because of the above-mentioned massive MIMO processing that will require the use of a very large number of radiating elements. Moreover, additional antennae compatible with new 5G bands will no doubt also need to be deployed. So the re-use of existing masts could very well be problematic, and new (possibly collocated) transmission sites will need to be found.

2. This search for new sites will also need to be carried out to install small cells in semi-elevated locations, but with extra care as the density of these installations will undoubtedly be high: operators will thus be required to deploy their equipment on urban furniture and infrastructures such as bus shelters, lampposts, public buildings, billboards, etc.

As a result, public authorities will need to keep a close eye on the matter and, if necessary, adopt measures that will facilitate 5G rollouts.

---

[42] http://www.lemag-numerique.com/wp-content/uploads/2015/10/WP_-Souverainete_Telecoms_PetitesCells_FINAL.pdf

[43] IFER = *Imposition forfaitaire sur les enterprises de réseaux*

3.3.3   5G networks' regional coverage and backhaul

Ensuring regional connectivity will be one of the challenges for this new generation of mobile networks.

The diversity of use-cases envisioned for future 5G networks, the geographical distribution of which is not yet precisely known, must be factored in when addressing coverage issues.

The higher frequency bands that are being put forth for future 5G networks, along with the potentially very substantial bandwidth consumed by these new uses, pose the question of the regional foothold of these networks. As a matter of fact, mobile networks have never before employed such high frequencies whose use will require a large number of relay antennae to be installed.

Moreover, connecting the 5G installations to the network will push to the fore the question of the cost of connecting them via optical fibre, which will no doubt be necessary in most instances to ensure the expected quality of service. The industry needs to design the technologies that will make it possible to minimise the cost of 5G rollouts in rural areas.

3.4 Net neutrality issues

European regulation on safeguarding an Open Internet[44], adopted by the European Parliament and Council on 25 November 2015, for which European regulators required an additional nine months to specify the rules governing its application, introduces the principle of net neutrality as one of the top priorities in the standardisation hierarchy.

Net neutrality is an overriding principle that guarantees equal treatment for all data traffic on the internet. In particular, it excludes any form of discrimination with respect to the source, the destination or the content of data flows.

On 30 August 2016, BEREC (the Body of European Regulators for Electronic Communications) published guidelines for national regulators on the enforcement of the European Open Internet regulation[45].

During the BEREC public consultation[46] on its draft net neutrality guidelines, several enterprises and electronic communications sector stakeholders took the opportunity to deliver a clear-cut point of view in their *"5G manifesto for timely deployment of 5G in Europe"*[47]. This manifesto aims at warning public authorities against a too restrictive approach to traffic management, and especially of the supposedly negative effects that, in their opinion, an overly strong enforcement of net neutrality could have on the 5G rollout roadmap.

---

[44] Regulation (EU) 2015/2120: http://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32015R2120&from=EN

[45] Internet service providers (ISP) can employ reasonable traffic management measures for certain categories of traffic, under non-discriminatory conditions and provided they are transparent, proportionate and justified by objective technical requirements, and not used to serve their own commercial interests. ISPs can also distinguish certain services, called specialised services, from their internet access service, without them affecting the latter, provided these services have specific, objective transit requirements.

[46] http://www.berec.europa.eu/

[47] http://telecoms.com/wp-content/blogs.dir/1/files/2016/07/5GManifestofortimelydeploymentof5GinEurope.pdf

BEREC's opinion[48], in response to the contributions to the aforementioned public consultation, recalls that the rules of net neutrality are technology neutral and therefore apply to 5G networks. Moreover, BEREC guidelines clarify that network slicing which networks will need to use to respond to the different use cases (mMTC, eMBB, uRLLC) planned for this new generation (cf. 1.4) system, may be used to deliver specialised services[49].

The subject of net neutrality as it applies to future 5G networks is still very much open and unexplored, but new analyses could be performed parallel to the work being done on defining 5G. Arcep is open to discussion and will also work, within BEREC, towards bringing the required clarity to the ecosystem.

---

[48] http://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/6161-berec-report-on-the-outcome-of-the-publi_0.pdf

[49] http://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/6160-berec-guidelines-on-the-implementation-b_0.pdf - footnote 26

Annexes

Annex 1   Definition and standardisation work

A new generation of mobile telephony is defined primarily by two bodies: ITU (International Telecommunication Union) and 3GPP (3rd Generation Partnership Project). These two, respectively public and private, organisations are dedicated to defining the objectives, standards and technical specifications of the new technology in question.

ITU

The first large-scale 5G commercial rollouts are expected to take place in 2020. As mentioned earlier, the exploratory phase – which provides an opportunity to determine demands and identify the most promising techniques and technologies for these future 5G networks – has already begun. Although a number of organisations and consortia are involved in defining 5G, 3GPP will very probably be the central standardisation body for its technical specifications.

Whatever the case may be, ITU (International Telecommunication Union) is vital to defining the technologies and standards that govern any new generation of IMT (International Mobile Telecommunications) at the global level.

These IMT standards are established with the involvement of public authorities and industry players, and have provided the framework for the evolution of mobile communication services around the world, since the beginning of IMT standardisation, with IMT-2000 (3G, UMTS), then IMT-Advanced (4G, LTE-A) and more recently IMT-2020 (5G).
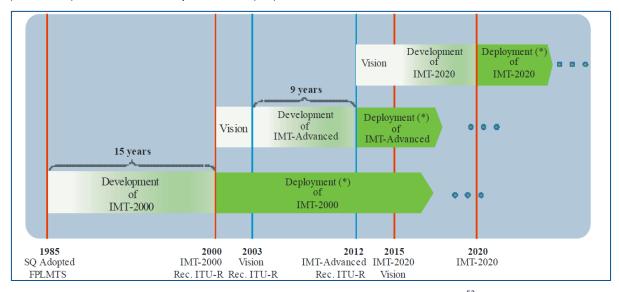
Figure 16. Timeline of IMT developments and deployments[52]

The advent of an IMT standard typically occurs in three main stages : a vision stage, a development stage and a deployment stage.

The vision stage – whose completion is marked by a document that is usually called a Vision Recommendation – is the stage during which ITU sets the framework and objectives for the future technology. The aim of this document is generally to define what the new technology will be in a more or less concrete fashion, what its characteristics will be, the uses it will enable, etc. Whether for

---

[52]Recommendation ITU-R M.2083-0 (09/2015), https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf

IMT-Advanced, the commonly used technical abbreviation attributed to the definition of 4G, or IMT-2020, the abbreviation attributed to definition of 5G, the vision stage lasts around three years.

Recommendation ITU-R M.2083-0 was published in September 2015, bringing to a close three years of work performed by ITU-R (Working Party 5D) on defining the framework and objectives for IMT for 2020 and beyond. It is this document that today serves as the basis of the different 5G research and standardisation work being done around the globe.

Next comes the standards development stage, based on the conclusions of the vision stage. Regarding IMT-2020, this development and standardisation work is already underway, and ITU plans on having completed it by 2020 to be able to satisfy the most pressing needs of the ITU members and organisations that want to deploy 5G as quickly as possible. This is all the more challenging as the development stage will last only five years, compared to 15 years for IMT-2000 and nine years for IMT-Advanced.

The different deadlines set for IMT-2020 within ITU can be found in Figure 17. The spectrum identification phases (indicated by the black triangles) coincide with the World Radiocommunications Conferences, of which the latest was WRC-2015 and the next is WRC-2019 (see 3.2).
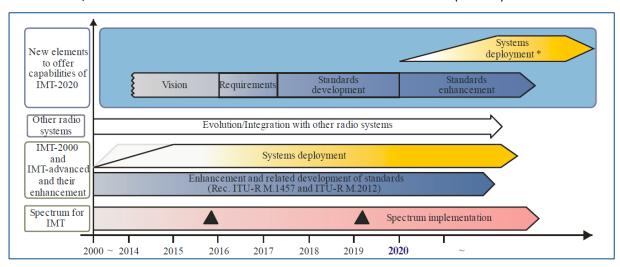
Figure 17. Stages and expected deadlines for IMT-2020[53]

All of the work that ITU is conducting on IMT-2020 is following a roadmap that is detailed in Figure 18, with the completed stages indicated in green and those to come in blue. The work that is underway today concerns 5G spectrum aspects, prerequisites and assessment criteria, along with the different technical studies and proposals, which are a prerequisite to the standardisation phase of the work.

---

[53]Recommendation ITU-R M.2083-0 (09/2015), https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf
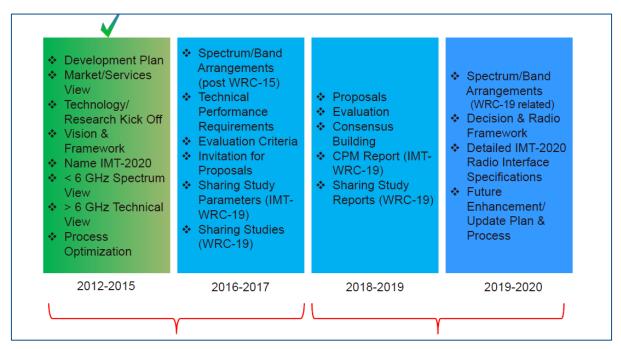
Figure 18. WP 5D[54] work programme

3GPP

Based on ITU Recommendations, 3GPP has played a major role in the success of LTE over the past several years, which has been the fastest growing cellular technology to date: never before has a new radio standard been adopted and deployed as rapidly and as widely from the finalisation of the first version of its standardisation (for LTE Release 8, in December 2008).

For the first time, under the LTE acronym, the entire mobile industry agreed on a single new technology (contrary to 3G where 3GPP and 3GPP2 co-exist, each backing their version of 3G standardisation that complies with IMT-2000 criteria), and so enabling unprecedented economies of scale and momentum in the ecosystem.

After Release 8, the work performed by 3GPP has been centred on the following strategic areas:

- Enhancing LTE radio standards to further improve capacity and performance;
- Enhancing system standards to make LTE and EPC (Evolved Packet Core, the core LTE network technology) available to new business segments;
- Introducing improvements for system robustness, especially for handing exponential smartphone traffic growth.

These areas of focus have made it possible to map out the general path of evolution from LTE to LTE-Advanced (Release*s* 10 to 12) and later LTE-Advanced Pro (Release*s* 13 and 14) while awaiting 5G.

The ambitiousness of 5G requirements, the tight timeline "imposed" by the market, along with the different national tendencies have pushed 3GPP to define two stages of specification work:

1. A first stage that will end in the second half of 2018, with the termination of Release 15, and which will address the most urgent matters with respect to commercial requirements;

---

[54] https://www.itu.int/en/membership/Documents/missions/GVA-mission-briefing-5G-28Sept2016.pdf

2. A second stage that will end in December 2019, which marks the end of Release 16, which will address the other use cases and prerequisites identified in the IMT-2020 Vision Recommendation.

The Work Plan for Release 15, the first set of 5G specifications, was ratified at the plenary meeting of the TSG#72 (Technical Specification Group #72) 3GPP working groups in June 2016. This plan includes a set of intermediate tasks and checkpoints to steer ongoing work in the different groups. This Work Plan sets out the transition from current studies to the standardisation phase of the work:

- Starting in December 2016: beginning of standardisation work by TSG SA2 (Technical Specification Group System Architecture);

- Starting in March 2017: specification begins for the 5G NR (New Radio) interface in the TSG RAN (Technical Specification Group Radio Access Network).
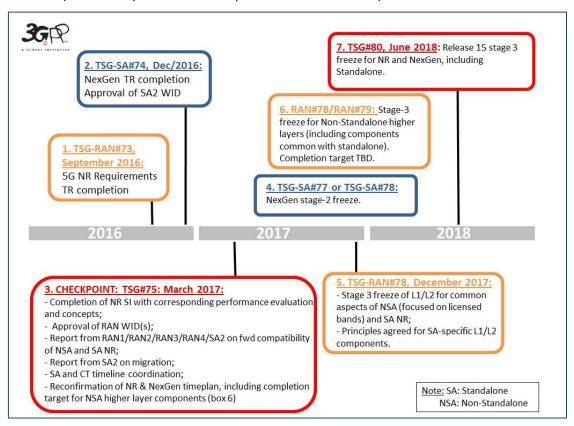
Figure 19. Roadmap for 3GPP standardisation groups[55]

If 5G is viewed as the technology capable of transforming society, and even of ushering in the next industrial revolution by impacting multiple (vertical) sectors with new business models and in a way that benefits the economy, notably in Europe, in all likelihood it will initially be driven forward by eMBB (Enhanced Mobile Broadband, cf. 1.2).

With this in mind, 3GPP defined the framework for Release 15, stage 1 specifications of 5G standardisation, as detailed earlier, whose work will focus on the following:

- Standalone and Non-Standalone NR operation: the Standalone version to work independently of LTE technology, and the Non-standalone NR being highly interoperable with LTE[56].

---

[55] https://5g-ppp.eu/wp-content/uploads/2016/11/01_10-Nov_Session-3_Dino-Flore.pdf

- Target use cases: starting with Enhanced Mobile Broadband (eMBB) as well as Low Latency and High Reliability to enable some URLLC (Ultra-reliable and Low Latency Communications, cf. 1.2) use cases.
- Simultaneous examination of frequency ranges below 6 GHz and above 6 GHz.

Release 16, the second phase of 5G specifications, will then focus more on other segments such as mMTC (Massive Machine Type Communications, cf. 1.2), for instance, to tackle IoT use cases that require higher quality of service than what LPWAN *(Low-Power Wide-Area Network)*[57] can provide.

---

[56] In its non-standalone version, the NR control plane is LTE's. in other words, the 4G network controls the 5G carriers and spreads users across the different bands and technologies.

[57] Sigfox and LoRa LPWAN

Annex 2   Entities Arcep met with

Annex 3   Figures

White Paper

Designing 5G-Ready Mobile Core Networks

Prepared by

Gabriel Brown
Senior Analyst, Heavy Reading
www.heavyreading.com

on behalf of

affirmed®

www.affirmednetworks.com

and

JUNIPEr®
NETWORKS

www.juniper.net

September 2016
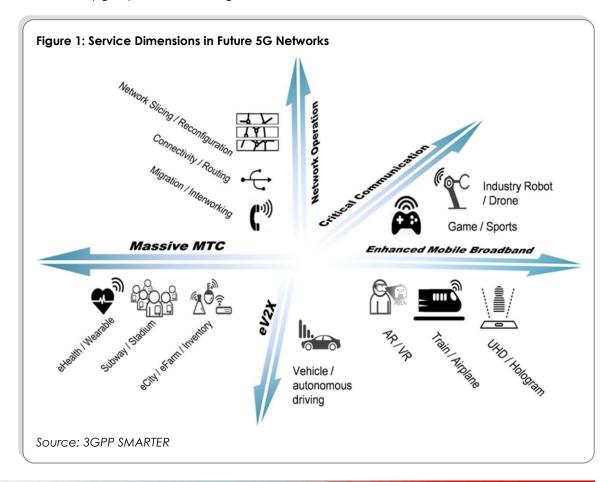
5G Core Service & Performance Requirements

An industry "Vision for 5G" is now established and supported worldwide. Services and performance requirements have been identified, and the industry is moving into the development and implementation phases with a view to commercial launch of 5G mobile service around 2019 and fixed wireless access using mmWave radio from 2017. To meet these schedules, leading operators – especially Tier 1s with high expectations and aggressive deployment timelines – are working to make their networks 5G-ready ahead of commercial deployment.

This white paper argues that with a 5G-ready technology strategy, operators can prepare for rapid 5G service launch in a way that optimizes their investment in next-generation IP and mobile core platforms over the next three years. Specifically, it discusses 1) how 5G services drive a requirement for an IP services fabric to connect the distributed data centers that will host 5G network functions, content and applications and 2) the development of cloud-native, service-orientated core networks for advanced 4G and 5G networks.

5G Services Dimensions

Development of the 5G system architecture, and associated core network, is being led by 5G service requirements. **Figure 1** shows the output from the 3GPP SMARTER study group that has investigated the service dimensions of 5G.

Figure 1: Service Dimensions in Future 5G Networks

Source: 3GPP SMARTER
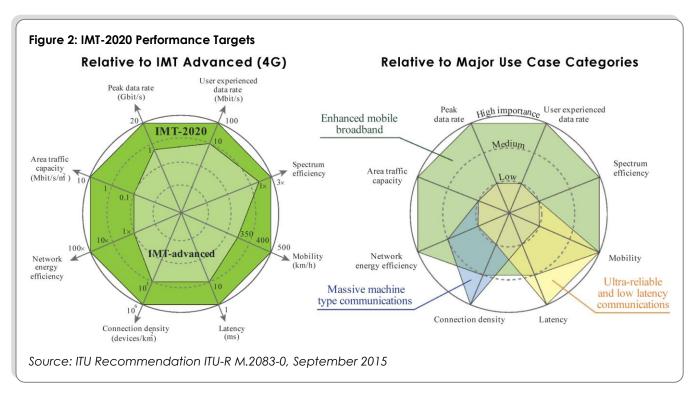
The chart maps closely to the now famous triangle representation of 5G services made up of enhanced mobile broadband (eMBB), low latency and mission critical communications and massive machine-type services (mMTC). However, it adds two important axes: a dedicated eV2X axis for vehicular services and, significantly, a network operations axis.

The dedicated operations axis is important because it makes explicit that automation is fundamental to 5G. To address new commercial opportunities, operators need the ability to create service-specific "network slices" that extend end to end across the infrastructure. Automating set up and management of these "network slices," including configuration of the underlying IP network, is critical to the 5G business case.

These service dimensions incorporate many different performance requirements. **Figure 2** shows IMT-2020 performance targets, developed by the ITU, for 5G. To the left, these are shown in comparison to IMT Advanced (4G). To the right, performance requirements are mapped to the three major use-case categories.

Figure 2: IMT-2020 Performance Targets

Source: ITU Recommendation ITU-R M.2083-0, September 2015

Probably the most influential factor on 5G physical network deployment will be the low-latency requirement. To deliver services over a wide-area network (WAN) with a consistent 5-10ms roundtrip time is extremely challenging and will drive a new system architecture deployed on a distributed cloud infrastructure.

A 5G-Ready Network Investment Strategy

Operators now have a good idea of what 5G will look like in scope and a reasonable view into the initial deployments models. Development of normative standards for 5G new radio (NR) and the next-generation core network (NG Core) is underway

with the first specifications – and the first "compatible" pilot deployments – expected in mid 2018. Although there is uncertainty about the details of the specifications under development, the industry has already undertaken significant R&D work on 5G candidate technologies, making it is possible for operators to pursue a 5G-ready network strategy with a reasonable degree of confidence that they will be able to rapidly deploy 5G, at scale, as the technology becomes commercially available.

From an IP and core network perspective, five tiers of a 5G-ready investment strategy are, from the ground up, as follows:

1. **Develop a distributed data center footprint**. The Central Office Re-architected as Data Center (CORD) model provides a good reference. The idea is for operators to develop physical assets located close to the edge and transform them into distributed data centers. The edge location is critical for low-latency 5G services and, as Mobile Edge Computing (MEC) shows, the edge-cloud is valuable in many networking scenarios – it is not solely a 5G investment. In time, the distributed data center will also come to support Cloud RAN (C-RAN) and Virtual RAN (V-RAN) hub sites and will host core network functions, such as distributed user-plane nodes.

2. **Create an "IP Services Fabric" for 5G with software-defined networking (SDN) control**. Edge-cloud locations running 5G network functions and services will require high-performance, secure connectivity. Investment in wide-area SDN is already underway to create an "IP services fabric" that connects centralized data centers, distributed cloud locations and cell sites. This IP services fabric provides routing, security, service chaining, redundancy and orchestration. Again, this investment is, in many aspects, not unique to 5G, but common to high-performance networking at the cloud edge.

3. **Invest in cloud-based Evolved Packet Core (EPC) and NG Core.** With a distributed cloud infrastructure, it is logical to redesign core network functions to take advantage of this asset to meet the strict latency requirements of 5G services and efficiently manage data traffic. Some aspects of NG Core are already reasonably well understood, (for example, control- and user- plane separation or "CUPS"), while others remain works in progress (for example, mobility and session management). In general, expect control-plane separation to apply across the 5G architecture, from radio to core.

4. **Use network slicing for a service-optimized core.** Network slicing is an important bridge from a 4G core to a new 5G core and is a key component of a 5G-ready network investment strategy. In this model, virtual network functions (VNFs) are created per service and data is routed via an optimized processing path across the network. By isolating services into virtual network slices, operators can offer better security, more efficient transport, optimized core network processing and appropriate service quality. A network slice created for an Internet of Things (IoT) customer, for example, may support both 4G and 5G access networks.

5. **Automation and service orchestration.** Automating network operations provides the ability to launch new services rapidly, to reach new market segments, to evaluate success/failure quickly, and then to modify, scale and repeat. Significant progress is being made in applying service orchestration to 4G-Long Term Evolution (LTE) networks, especially where software-based core networks are deployed. Network slicing is a good example of the need for automation because there will be many more logical entities to commission, provision and manage. "Automated service orchestration" is a good, tangible example of 5G-ready functionality.
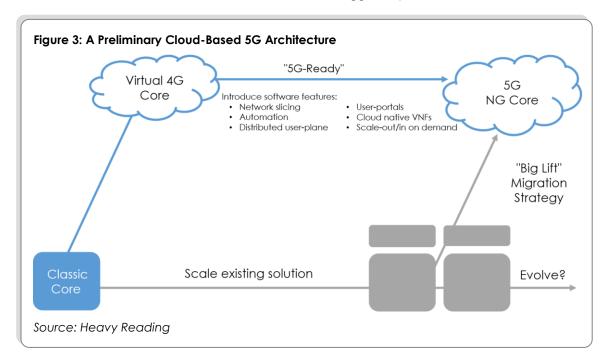
The 5G Core Network

The new 5G core network is now reasonably well understood: It will be "cloud-native," it will make extensive use of network slicing, and it will operate in concert with a new model-driven service orchestration layer. The industry can therefore start to develop NG Core products that can be adapted as standards emerge and solidify.

From a practical perspective, for the initial deployment, operators can support 5G radio access on a "5G-ready" EPC and then migrate to a new NG Core over time. In both cases, the "IP services fabric" provides SDN-controlled connectivity and related IP services, across a distributed cloud infrastructure.

Virtualized & Cloud-Native Mobile Core

Operators have made good progress on virtual EPC over the past couple of years. The largest networks now support more than 15 million subscribers (AT&T has discussed this publicly); some progressive operators have started to refresh their main EPC networks using multivendor network functions virtualization (NFV) (e.g., Docomo, Etisalat); and others (e.g., Vodafone) have deployed virtual core elements for IoT services.

Figure 3 shows two migration paths to a "cloud-native" 5G core. The blue line shows a virtualized 4G core as a stepping stone to a "5G-ready" core and then a full 5G core. The grey line shows a more conservative option where the operator scales the classic EPC in the near-term and then makes a bigger leap to 5G later.

Figure 3: A Preliminary Cloud-Based 5G Architecture

Source: Heavy Reading
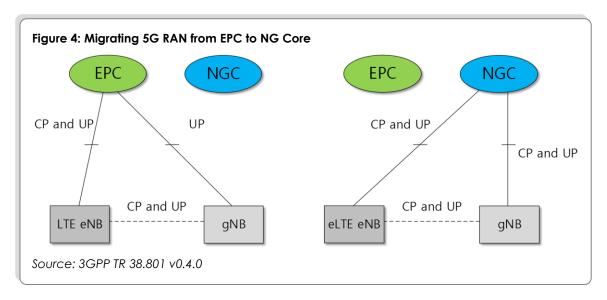
There is debate about what exactly "cloud native" means, particularly in the context of stateful telecom network functions. Nevertheless, Heavy Reading's established view is that operators that deploy virtual EPC – even at small scale – gain experience in how these systems work, how to operate them, and how to evolve them, and that in turn will give them a sustainable advantage in the longer-term transition to software-centric 5G mobile networks.

Migrating 5G RAN From EPC to NG Core

A 5G-ready core strategy is determined, in part, by how the operator plans to introduce 5G radio. There are two basic possibilities: operate 5G RAN using an EPC or using a new NG Core. In practice, operators with plans to launch 5G early are likely to start with an EPC and migrate to NG Core over time, as shown in **Figure 4** below.

Figure 4: Migrating 5G RAN from EPC to NG Core

Source: 3GPP TR 38.801 v0.4.0

The diagram is taken from the most recent version of the 3GPP "Study on New Radio Access Technology" (TR 38.801), which will inform development of standards in Release 15. Several permutations of the architecture are under consideration; however, in simple terms, to the left, the new 5G base station (gNB) user-plane interface connects directly to a 4G EPC, while the control-plane functions, such as tracking, paging, etc., are provided by an evolved 4G base station (eNB), which in turn also connects to the EPC. In this scenario, the EPC requires little or no modification, making this a fast and simple way to deploy 5G radio from a core perspective.

Over time, both 4G and 5G base stations can migrate to a new NG Core, which will provide both control- and user-plane functions. At this stage, NG Core becomes the primary core network for 4G and 5G access, as shown to the right in the diagram. This is conceptually similar to how EPC supports 3G and 4G access networks.

Note that in cases where 5G radio is deployed for fixed wireless access, there is no need for an LTE RAN to provide control-plane functions to the 5G user device; a standard EPC is sufficient, although it would need to provide session management for the 5G access. Some of the first 5G deployments are expected to use this model.

Network Slicing: A Key Bridge to 5G

Network slicing is one of the key bridges between the 4G and 5G core. To support diverse service types, operators will use multiple core networks deployed as "network slices" on a common IP services infrastructure. The idea, shown in **Figure 5**, is to create virtual core network instances (or "slices") dedicated to different services. Each slice can be optimized for the traffic profile and the commercial context of the associated service – for example, IoT, public safety, mobile virtual network operator (MVNO), connected car, voice over WiFi or enterprise services. Network slices can be two dimensional in the sense that they can be both service- and customer-specific.

Figure 5: Dedicated Packet Core Network Slices

- Virtualization enables multi-tenant and single-tenant private core networks
- "Slices" configured according to service type & traffic profile
- Multiple options to steer user traffic into a core network processing path
- Mechanisms in 3G/4G include APN routing, MoCN, DECOR

Source: Heavy Reading

IoT is an example of how operators can use "slicing" to support a 5G-ready core strategy. Since many of the core network design parameters are similar between 4G and 5G, investment in a software-based IoT core network can be made with the expectation that the same core (with software updates) will also support IoT services on 5G in future. Moreover, because devices from both access types will connect to a common IoT core, operators will be able to develop integrated 4G/5G IoT strategies that optimize investment and enable them to go to market with narrowband IoT services before 5G specifications are released and radio equipment is deployed.
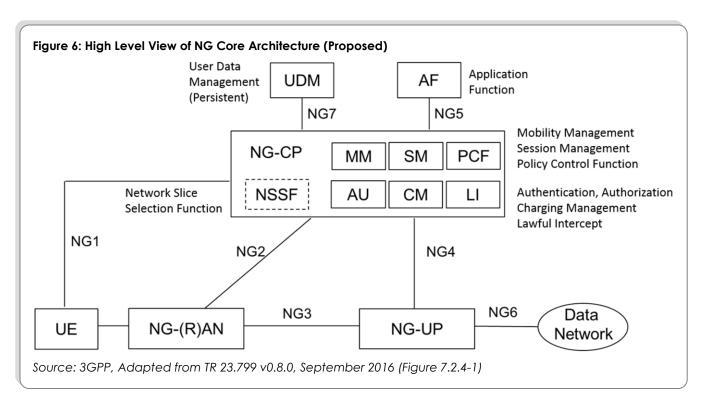
5G will, naturally, bring more capabilities to network slicing. Most importantly, it is expected that the slice will run end to end across the RAN, core and transport network. Radio is typically deployed as a shared resource, which means that 5G slicing will involve advanced self-organizing network (SON) capabilities and hierarchical, slice-aware scheduling on the air interface. On the network side, the operator can use SDN to reserve resources for the network slice on the IP services fabric.

NG Core for 5G

A new core network offers important benefits, particularly for services with demanding performance requirements, and is an important part of the 5G architecture. The high-level view of the NG core network architecture, as it looks in September 2016, according to TR 23.799 (the Technical Report on the NextGen System Architecture) is shown in **Figure 6**.

The chart shows the primary elements of the new 5G system architecture, including the device (UE), the radio access network (NG-RAN) and the core network that comprises the control-plane (NG-CP) and user-plane (NG-UP) functions. The separation of control and user planes is a direct extension of the "CUPS" concept being developed for advanced EPCs in 4G networks and is another example of how operators can invest in 5G-ready core network designs in the near term.

Figure 6: High Level View of NG Core Architecture (Proposed)

Source: 3GPP, Adapted from TR 23.799 v0.8.0, September 2016 (Figure 7.2.4-1)

It is interesting to compare this new functional architecture with the existing EPC. **Figure 7** maps the new NG Core elements to their equivalent in EPC. It shows significant overlap between the EPC and NG Core and provides confidence that today's state-of-the-art EPC is reasonably close to the next-generation core in terms of functionality, albeit that the interfaces and protocols will change/evolve.

Figure 7: Mapping NG Core & EPC Functions

NG Core Network Function	Approximate Equivalent EPC Function
UDM (User data management)	SPR
AF (Application function)	AF
NSSF (Network slice selection function)	MME (DECOR) + HSS
MM (Mobility management)	MME + SGW / MME + SGW-C
SM (Session management)	PGW/PGW-C
PCF (Policy control function)	PCRF
AU (Authentication & Authorization)	HSS/AAA
CM (Charging management)	PGW + PCRF, OCS, OFCS
LI (Lawful intercept)	LI
NG-UP (User-plane function)	SGW + PGW / SGW-U + PGW-U

Source: Heavy Reading, Affirmed, Juniper

Key Capabilities of the Next-Gen System Architecture

In addition to proposed architectures, the Technical Report provides guidance on the features and services NG Core should support. It identifies 19 key issues for the Next-Gen System Architecture, which can be grouped into four categories:

- **Flexible Deployment**: Including network slicing, user-plane network selection, network function granularity (e.g., decomposition of VNFs and service chains), interworking and migration, a policy and charging framework.

- **Flexible Access Support**: Including variable core/access splits, a flexible authentication framework (with varying credentials according to device-type, use case and policy), support for relays and multi-hopping, improved network discovery and selection mechanisms.

- **Connectivity**: Including session management, mobility management, a quality-of-service (QoS) framework and service and session continuity (e.g., across accesses).

- **Adapting Existing Capabilities**: Including network/service capability exposure, multicast and broadcast, IP Multimedia Subsystem (IMS) support, off-network communications (e.g., *ad hoc* networks for public safety).

Implementation of these capabilities is tightly linked to both the design of the NG Core and the underlying IP service infrastructure. The 3GPP specifications do not reference connectivity services, but, in practice, there is a close relationship.
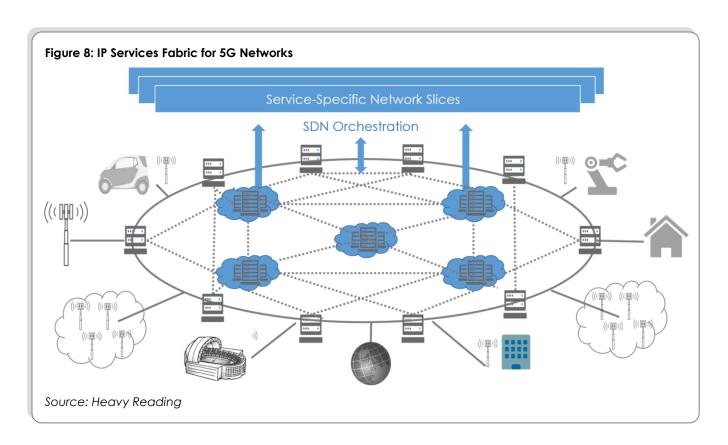
A Distributed 5G Services Fabric

The infrastructure onto which 5G will be deployed should support multiple, demanding use cases. This drives a need for a high-performance wide-area IP services fabric controlled (or "orchestrated") by SDN. This IP services fabric should provide connectivity between many different distributed data centers in a meshed architecture that provides resiliency and scalability, and should be programmable such that it can support dynamic, service-specific network slices. In effect, the IP services fabric makes distributed centers act in a unified manner – i.e., behave as one integrated data center. The concept is shown in **Figure 8**.

The design of the distributed IP services fabric is formed by the variety of use cases, the requirement for low latency and high availability, and the need to scale efficiently. In tandem, the mobile network architecture will evolve to a more distributed model. To meet 5G service requirements, NG Core will be deployed using the "CUPS" model, with user-plane nodes hosted at distributed data center locations, and control-plane nodes at more centralized locations, interconnected by the IP services fabric.

Service orchestration will generally be domain specific, with SDN-controlled IP services, cloud resource management, NFV lifecycle management and end-user service orchestration, all operating quasi-independently in a layered architecture. Coordination between the layers will use "cross domain orchestration" to ensure a network slice contains all the networking components needed to deliver the service.

Depending on how the architecture evolves, there will be a need to support highly accurate, 1588v2-based timing (frequency and phase) within this IP services fabric. Virtual RAN/Cloud RAN is an example on the network side that benefits from accurate network timing. On the customer side, it is expected that some market segments, such as the financial industry, will also require timing to be incorporated into network slices.

Figure 8: IP Services Fabric for 5G Networks

Service-Specific Network Slices

SDN Orchestration

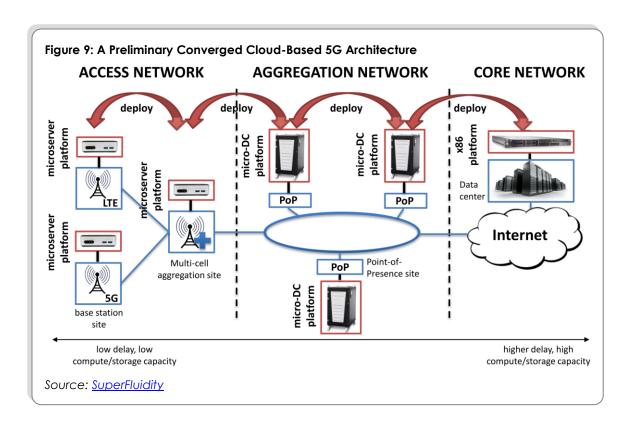*Source: Heavy Reading*

Distributed Data Centers

To provide access network services, telecom operators have a physical footprint – in the form of central offices, base station controller sites and transport aggregation sites – that they can convert into distributed or micro data centers.

Depending on the service requirements, these data centers can be used to terminate access connections using virtualized edge functions and become the obvious place to host latency-sensitive content and applications. For centralized cloud providers, this capability is harder to replicate unless they build out an extensive physical footprint or partner with an access provider, and, therefore, this strategy provides operators with a potentially important and sustainable competitive advantage.

The "edge cloud" model will be critical to 5G and is the subject of detailed R&D across the industry. As an example, a preliminary architecture developed by the SuperFluidity project (a European 5G research project) is shown in **Figure 9**. Work on this particular project, which is academic in focus but contains several industrial partners, is scheduled to finish by the end of 2017.

As the name implies, the project aims to achieve superfluidity in the network: the ability to instantiate services on the fly, run them anywhere in the network (core, aggregation, edge) and shift them transparently to different locations. This is the essence of the "IP Services Fabric" for 5G. In this new network, operations will also need to move from a "red light, take action" model to an automated model where the orchestration and associated next-gen operations support system (OSS) detect and fix problems with the IP and NFV layers, and re-optimize the network accordingly.

Figure 9: A Preliminary Converged Cloud-Based 5G Architecture

Source: SuperFluidity

Central Office Re-architected as a Data Center (CORD)

There are several industry and service provider initiatives to define the architecture for SDN-enabled distributed data centers. The Central Office Re-architected as a Data Center (CORD) initiative hosted by the Linux Foundation is one example of how operators aim to make better use of these assets.

According to the CORD initiative, AT&T alone operates 4,000 to 5,000 central offices, each serving 10,000 to 100,000 residential, enterprise and mobile customers. These central offices contain fragmented vendor hardware with multiple physical appliances installed per site (AT&T said it has 300+ unique appliances deployed in its central offices nationwide). The opportunity is to re-architect these central offices to support edge cloud infrastructure and deploy VNFs in place of appliances.

Each CORD location will be connected using an SD-WAN, making it possible to load-balance content and NFV workloads across the distributed cloud using the same SuperFluidity concepts discussed above. In the CORD case, the focus is on the ONOS controller, but in the sense that this is a generic architecture, multiple SDN controller options are viable and attractive. Note that for I/O-intensive workloads, the selection of an edge location should consider the physical transport resources available at the site. It would not make sense, for example, to deploy a 5G user-plane node for fixed wireless access in a central office that does not have a high bandwidth connection to the Internet.

In March 2016, the Linux Foundation announced the M-CORD initiative for mobile operators, backed by AT&T, SK Telecom, Verizon, China Unicom and NTT Communications. It highlighted three key aspects of the architecture, saying that it:

- Will use the same CORD principles of elastic commodity cloud and SDN to bring data center economies and cloud agility to the mobile edge.

- Will demonstrate integration of disaggregated/virtualized RAN, disaggregated/virtualized EPC and mobile edge services.

- Will partner with the SDN controller groups to accelerate adoption of open source SDN and NFV solutions and realize the benefits of the cloud.

Given the timing of the work, we expect M-CORD and similar initiatives to turn their focus toward "5G-ready" core networks as the architecture, interfaces and protocols for NG Core become more clear. A common CORD and M-CORD implementation, with the same architecture and foundational technologies, will create a good foundation for fixed-mobile network convergence, enabling access agnostic services – an important objective of many operators pursuing 5G.

Mapping Mobile Core to Distributed Cloud

The NG Core for 5G and EPC for advanced 4G networks must be mapped to the distributed cloud architecture. One approach would be to simply deploy more packet gateways (and mobility controllers) at the edge of the network to meet capacity and performance demands.

The challenge with this is that today's centralized packet core deployments are characterized by complex integration with surrounding network functions, such as policy, charging, IMS, SGi-LAN and routing services. By moving this model to the edge, the operator would, in effect, have to "distribute complexity," which is costly to deploy and, in particular, to manage. To meet 5G performance, scalability and automation requirements, a new architecture for packet core is needed that will make operation in the edge cloud infrastructure simpler and faster.

There are many aspects to this new architecture. Part of the solution is CUPS, as is currently being developed for 4G-LTE core networks. This involves extracting the control-plane functions from the gateway to leave a simpler, user-plane node. The gateway thus is "split" into S/PGW-U and S/PGW-C components that can that can scale independently, as shown in **Figure 6** above. A key benefit of the architecture is that the control plane, and all the associated complex interactions, can be centralized, while the user plane is distributed across the IP services fabric and scaled as required by the traffic load. This is shown in **Figure 10**.
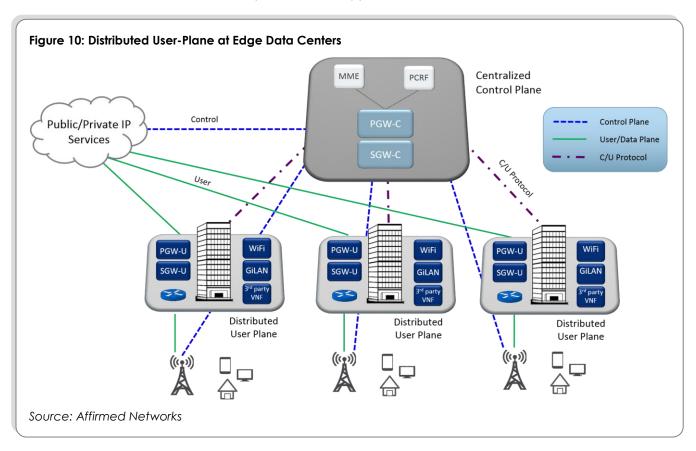
Depending on the scaling needs, the S/PGW-U functions deployed at the edge data center can be implemented in several different ways: on a router, on a white box switch (potentially), on an existing PGW platform, as a virtualized function, or as part of a vRouter. Virtualized user-plane nodes can more easily be placed at the optimal location, as determined by the use case and this flexibility is a strong argument to deploy S/PGW-U nodes as VNFs rather than as hardware functions.

The S/PGW-C components would similarly be deployed as virtualized functions on a cloud platform, typically at a more centralized location. There may be an opportunity to collapse MME, SGW-C, PGW-C functions and perhaps other 3GPP control-plane functions into some form of unified "mobility controller" node. This has the potential to simplify operations and this same model is being pursued for NG Core.

In this architecture, virtualized SGi service LAN components can be placed where appropriate for the traffic coming from the distributed user-plane nodes. For example, some SGi functions could be on router-based compute blades, or on COTS x86

servers deployed at the distributed site, while others could be in the central data center or close to Internet peering points. The selection of the site for the SGi functions is carried out by the central orchestration system, which steers traffic into service chains on a per subscriber, application, bearer, device or combination basis.

Figure 10: Distributed User-Plane at Edge Data Centers

Source: Affirmed Networks

This model of discrete control- and user-plane functions is expected to be fundamental to 5G and the NG Core. In this sense, CUPS can be viewed as an important part of a 5G-ready investment strategy.

A 5G-Ready Core Is Now a Priority

The industry now has a reasonable view of 5G service requirements, and progressive operators with aggressive deployment timelines are now working to prepare their networks for rapid deployment of 5G when equipment is available. This is driving investment in the critical IP services network needed to connect the edge cloud locations that will run 5G network functions, content and services.

On the mobile core side, development of cloud-native, service-orientated core networks for advanced 4G and 5G networks is underway. Network slicing provides a conceptual bridge between 4G and 5G investment and facilitates a faster insertion of new services into the network. Similarly, CUPS provides a reference for NG Core and the new 5G network architecture. With a 5G-ready technology strategy, operators can prepare for 5G service launch in a way that optimizes their investment in next-generation IP and mobile core platforms over the next three years.

About Affirmed Networks

Affirmed Networks' NFV solution has become the standard for the world's top mobile operators, who are embracing new business models and building new revenue streams by making the transition to virtualized architectures. The company's technology portfolio includes the Affirmed Mobile Content Cloud, the Affirmed Wi-Fi Gateway (serving as a TWAG/TWAP and an ePDG), Affirmed Service Automation Platform (ASAP) and Affirmed Virtual Probe and Analytics Solution. These virtualized solutions have come to represent the present and the future of virtualized mobile networks with extreme scalability, remarkable flexibility, comprehensive network orchestration and future-proof solutions for a 5G-ready architecture. Please find more information at www.affirmednetworks.com.

About Juniper Networks

Juniper Networks (NYSE: JNPR) is in the business of network innovation. From devices to data centers, from consumers to cloud providers, Juniper Networks delivers the software, silicon and systems that transform the experience and economics of networking. The company serves customers and partners worldwide. Additional information can be found at www.juniper.net, or connect with Juniper on Twitter and Facebook.

Trending | New Products | Power | Fire, Life Safety | Electrical, Lighting | Building Types | Codes and Standards | HVAC

Energy harvesting power for the Internet of Things

Energy harvesting wireless technology easily interconnects thousands of individual devices in a system, opening up unlimited processing and monitoring applications.

Laurent Giai-Miniet, EnOcean, Oberhaching (Munich), Germany
09/30/2013

Share

Energy is everywhere within reach; it just needs to be harvested. This is the principle of energy harvesting. Today, energy harvesting wireless solutions are already well established in the commercial building automation sector. But the technology is just getting started. New application fields for batteryless, wireless communication will be found to further enhance the world around us.

Based on energy harvesting wireless technology, a wide range of energy-autonomous applications are currently available for connected buildings that use motion, light, or temperature differences as their energy source such as batteryless switches; intelligent window handles; temperature, moisture, and light sensors; presence detectors, heating valves; and smart home systems. However, building automation is by no means where energy harvesting wireless ends.

Multiple interconnections

Everybody's talking about the Internet of Things (IoT). But how should billions of communicating devices be powered? The answer is by energy harvesting, and the reason is simple: Liberating sensors from external power, making them energy-autonomous, opens up unlimited processing and monitoring applications where cables or batteries represent an insurmountable hurdle. These features make energy harvesting wireless technology the ideal communication standard to easily and reliably interconnect thousands of individual devices in a system, as well as network them with other wireless protocols.

Today, energy harvesting wireless technology is widespread, providing M2M solutions in the building automation sector and bridging the control of light, HVAC, and other fields of building technology to smart buildings, smart metering, and energy management systems. This is the starting point to actuate further applications that lead to the IoT in the long term. The following four categories show what this could look like.

Monitoring and control

Wireless and batteryless technology significantly eases energy monitoring and control in buildings with little intervention into the existing systems. The wireless devices are highly flexible to install so that individual components, wall switches, sensors, and relay receivers can be easily networked to form an intelligent system without complex cabling. In addition, dispensing with batteries eliminates the burden of maintaining the devices' energy supply in a regular time period, which can be up to each year.

An example for such a flexible automation system is HVAC control. Here, a thermostat, VAV (variable air volume), or fan coil controller receives information related to occupancy, temperature, humidity, window position, or CO_2 from the respective batteryless sensors and controls the opening and closing of valve actuators for radiators, or dampers for VAV systems. At the same time, the controller sends status information to a central building automation system, and receives control messages from the BAS. This enables the building to be monitored from a central location that can be remote from the building itself, and building-wide settings, such as holiday shutdown, to be implemented. Enormous progress is also being made on the product side, leveraging advancements in energy harvesting. Self-powered radiator valves generate energy from the difference in temperature between the hot water and the surrounding air. This energy powers the communication with a controller or BAS system, and turns the valve itself. Without cables or batteries, these wireless devices are especially easy to install, and they require no maintenance.

In further optimized systems, central equipment such as boilers or air handling units are integrated into the wireless communication system, enabling scalable HVAC generation, visible and controllable over the Internet on a PC, tablet, or

smartphone.

Performing tasks

Alarm systems are a second field that batteryless wireless technology is opening up, due to its specific features. Here, the reliability requirements are much more stringent than those required for lighting controls. A system failure not only means a malfunction but also can cause much more serious consequences for other systems that depend upon the equipment being monitored. It's a fact that more malfunctions are caused by battery failures than by the electronics, especially in large systems. Energy harvesting overcomes this issue.

There are already various batteryless wireless water detectors available that use miniaturized solar cells or motion energy converters to power wireless signals that report water leaks in areas such as water supply networks in spacious industrial facilities. In the AFRISO universal module, for example, the EnOcean wireless signal immediately sends the leakage information to a gateway controller or directly to a valve, causing the main water pipeline or the affected supply line to be shut off. A notification is sent to the user's smartphone or smartpad at the same time to inform the user about the incident. In addition, the water valve can be opened and closed, independent of leakage notifications, by GSM connection via smartphone or smartpad.

Embedded processing

A major requirement of today's and the future energy supply is the Smart Grid. It's intended to network centralized and decentralized energy suppliers, including private homes producing electricity by photovoltaic installations, to an intelligent system that provides energy only when needed, updating in real time. This requires continuous data flow and processing from all involved parties, which means from millions of information points.

A key component is smart metering systems. To work reliably and cost-efficiently, interoperability between the meters is supplied by different manufacturers; this is why smart metering calls for standardized technologies. Consequently, the members of the EnOcean Alliance have defined a specific device communication protocol, the Automated Meter Reading (AMR) profile for batteryless wireless devices. Smart meter systems based on this open protocol are already available from a number of manufacturers. For example, Eltako meter components read and transmit the current electricity, water, and gas consumption, including accumulated meter figures, by means of energy harvesting wireless technology located at a variety of points inside a building. BSC software monitors and displays the current meter readings and compares them against default values. This makes all relevant data available for systems processing for intelligent energy management on demand.

Bridge to the cloud

Via similar gateways, the standard-based energy harvesting technology can also communicate with Ethernet, Wi-Fi, GSM/UMTS/CDMA, and other networks for integration in cloud services. Here, all data collected by batteryless wireless sensors is encrypted and transmitted to a cloud service over the Internet. The gateways connected to a control and visualization software by TCP/IP that can be used to control all relay receivers and sensors bidirectionally. Some manufacturers have developed a cloud solution that offers energy management as a service. Therefore, facility managers, building owners, and businesses can monitor important inventory, equipment, assets, and energy-related information from anywhere at any time, via the cloud. Critical building-related data is automatically pushed to the cloud, freeing owners and managers from the often-challenging coordination and expense of hosting on-site servers.

One of the major advantages of such a cloud-based solution is that the management system arrives completely precommissioned from the manufacturer and ongoing device commissioning is expertly done on behalf of the client and pushed out from the cloud. The users are granted unlimited access to their remote, dedicated virtual server with their own IP address, accessible from a desktop or smartphone — the perfect precondition for a deeply connected world of an IoT.

The energy harvesting market is growing and multiplies on a year-by-year basis. Forecasts show that this trend will continue, especially as the next generation of energy harvesting wireless solutions is just around the corner to pave the way to the Internet of Things.

Laurent Giai-Miniet is CEO of energy harvesting wireless solutions provider EnOcean and previously spent 20 years with Texas Instruments (TI), where he held several management positions including General Manager for Low Power RF Products (LPRF).

Related News:
- Energy harvesting for Internet of things - 02.07.2014 01:48
- Consumer market will drive industrial Internet of things for manufacturing - 06.03.2014 11:43
- How the Internet of Things will affect consulting engineers - 02.01.2014 11:35
- IP-all-the-way is the way to go - 23.07.2013 01:28
- Integration: Lighting and HVAC systems - 16.04.2013 01:30